Evolution and development of sensory systems in Chelicerata

By

Guilherme Gainett Cardoso Martins de Carvalho Florez

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Integrative Biology)

at the
University of Wisconsin-Madison
2023

Date of final oral examination: 19 May 2023

The dissertation is approved by the following members of the Final Oral Committee:

Prashant P. Sharma, Associate Professor, Integrative Biology

Marry C. Halloran, Professor, Department of Neuroscience

Jeff D. Hardin, Professor, Integrative Biology

Phillip A. Newmark, Professor, Integrative Biology

Gonzalo Giribet, Professor, Harvard University

Table of Contents

Acknowledgements

A Ph.D. is a long journey.

I traveled from Brazil to Wisconsin in late August 2017 to start my doctoral studies in evolutionary developmental biology of arachnids. It was a drastic change, both culturally and thermally. I vividly remember my first day in Madison, arriving at Hypatia Coop, a community house where I would live for many years. I was certain I had made a great choice, a mood shortly preceded by feeling clueless, and soon followed by thankfulness for the great opportunity ahead of me. Fast forward to 2023, emerging in a post-pandemic world, I feel happy and fortunate to have conducted this work. I have several people and institutions to thank for supporting me in so many ways over these six years.

None of this doctoral work would have happened had I not met, about a decade ago, a then a graduate student named Prashant. I met Prashant around my sophomore year in Brazil, when I had a very special chance to be an undergraduate research intern with Gonzalo Giribet (Prashant's Ph.D. advisor), to image harvestmen legs. Prashant was very supportive of my project at the time. I also believe the first embryo I've ever seen was a spider embryo that he showed me in a stereomicroscope, although I did not understand much of what he was doing with them at the time. Nonetheless, the fact he wasn't bothered by my constant questioning about the whereabouts of harvestman museum specimens and asking, "is this male, or a female?", with hindsight, said a lot about him as a future mentor. I had no idea that in a few years he would become my PhD supervisor. Prashant is one of the most enthusiastic, insightful, and creative scientists I know. It is still hard to believe the amount of time and resources he devoted to my doctoral training, and the immense confidence invested in me by supporting absolutely every experiment, fieldwork, and idea I proposed. Prashant taught me close to everything I know about molecular biology, from pipetting to troubleshooting in situ

hybridizations (a technique to visualize gene expression), cloning genes and manipulating embryos. I really appreciate the hands-on approach of his mentorship. Prashant also has a special gift to see the bright side of all the failed experiments and see the bigger picture of the findings. His uplifting comments were also essential to keep me motivated along this journey. My most sincere and heartfelt thank you, Prashant!

I also owe a lot of my benchwork training, successful cloning and PCRs, late-night experiments, and motivation to my lab mate Emily Setton. Emily has been an amazing friend throughout these years, making sure we always had snacks in the lab, music and fun facts about horses. I have yet to see someone as bright and hard-working as Emily. Her work is also present throughout this dissertation in so many ways. Thank you, Em!

The Sharma lab has been a place of many exceptional researchers, which have shared with me their great stories, lessons, joy for science, and contributions to all this work. My many thanks to my lab mates Andrew Ontano and Jojo Sardina, undergraduate researchers Audrey Crawford, Calvin So, Leonardo Barolo, Jakob Zehms, Keyton Friske, Kevin Corbett, John Zern, Ben Klementz (now a graduate student at the lab), Pola Blaszczyk, Hugh Steiner, Grace Hareid, postdocs Carlos Eduardo Santibáñez-López, Jesus Ballesteros (Chuy), Caitlin Baker, Luciano Palmieri, and Siddharth Kulkarni.

A big thank you to my committee members Gonzalo Giribet, Mary Halloran, Jeff Hardin, and Phil Newmark for their commitment and guidance during all the steps of this research journey. Being a committee member is a huge commitment, and so it is very generous of them to participate. A special thanks to Gonzalo. Meeting Gonzalo as an undergraduate was also a turning point in my career, where I could see how amazing the study of biodiversity could be and that harvestmen were indeed very special creatures. The opportunities and inspiration he provided me brought me where I am today. Thank you for your continuous support!

Backtracking a bit, it really all started at the Laboratory of Sensory Ecology and Behavior of Arthropods (LESCA) at the University of São Paulo, in Brazil, where I first heard about science and made friends for a lifetime. Rodrigo Willemart, my bachelor's and master's advisor, is a phenomenal mentor that opened all the doors for me, taught me so much, and introduced me to Gonzalo and Prashant. Rodrigo continues to give me phenomenal advice in science and life and has become a close friend. Thanks, Ro! Also in this lab, I met Gabriel Murayama, which despite the distance continues to be one of my closest friends. Thank you, Gabs, for always being there for me and being who you are! Traveling with these two has become one of the most enjoyable things in my life. Thank you for sticking with me as amazing friends and collaborators despite the long distance. Go Bear Hunters!

One of the most amazing experiences I had during this Ph.D. was conducting fieldwork across Israeli caves and deserts. This experience was truly priceless, both because of the amazing places and sceneries I have seen, and because of the new friends and colleagues I made there. One shall never forget about breaking the log, fighter jets, guano, hyraxes, giant camel spiders, inadvertently opening eyes under the dead sea, the stunning Sede Boqer desert, and amazing food and hospitality at Efrat's house. I will bring back Chuy here, without whom none of this field work would have been possible. Thank you for bringing me along and being an awesome travel mate! Many thanks to Efrat Gavish-Regev, for being the most amazing host one could hope for, and a great mentor along all the Israeli cave spider projects. Thank you, Shlomi, for your amazing hospitality and friendship, and being a real-life superhero rappelling in caves and guiding us through the desert. I am forever in debt with everyone in the Gavish-Regev lab and associated colleagues. Thank you, Igor Armiach, Zeana Ganem, Yoram Zvik, Sharon Warburg, Evgeniia Propistsova, Yuval Zaltz, Yael Lubin, Yotam Regev, Segula and Jimmy.

Collecting camel spiders in Las Salinas Grande in Argentina was also an amazing experience. True, the salt desert and animals are stunning, but the real thrill was meeting such a special research group as LABRE. My many thanks to all the new friends in Córdoba, Alfredo Peretti, Catalina Simian, Hernán Iuri, Fedra Bolatti, Franco Cargnelutti, David Vrech, Lucia Rosa, Paola Olivero, Mariela Oviedo, Rocio Pietri, Julieta Carbonari, Matias Izquierdo, and Camilo Mattoni.

Visiting the Marine Biological Laboratories at Woods Hole and taking the Woods Hole Embryology course in 2021 was a definitive turning point in this dissertation. The experiences I had there and the techniques I learned in the course made possible several discoveries reported here. It was there that I learned how to do multiplexed in situ hybridization with HCR, without which late stages embryos of daddy-longlegs could not be studied for gene expression. I received great career advice and meet wonderful people in my cohort of students. Thank you, Embryo21 friends! You made this experience truly amazing! Thank you, MBL, Nipam Patel and Heather Bruce, for being so supportive and willing to collaborate with me.

Many thanks to all the amazing staff at the Department of Integrative Biology (UW-Madison) that kindly assisted me through documents, grants, and life in Birge Hall. Thank you, Julie Larson, Kelin Boldis, Nazan, James Vandenberg, Michelle Kirch, Carol Cooley, Duane Barnes, Mike Malone, Brandon Dowd, Ronald, and Jeff. Thank you to iBio GO and all my fellow graduate students in the department!

Thank you, Ricardo Pinto-da-Rocha, for so much help and guidance in my time at the University of São Paulo. Thank you, Federico Brown-Almeida, for always being a point of support at the University of São Paulo and helping with reagents for fieldtrips in Brazil.

I really value the conversations about science, career advice, and guidance concerning postdocs that I receive from numerous people. Thank you, Marcos Simões-Costa, Tatiane Kanno, Ehab Abouheif, Abdou Khila, Bruno Veluttini, Bruno Medeiros, Tauana Junqueira,

Dave Shearwood, Leslie Babonis, Dede Lyons, Vanessa Barone, Juni Rajakumar, and many others also cited above.

Thank you, Sarah Swanson, the Newcomb Imaging Center at the Botany Department of UW-Madison, for all the amazing support with microscopy, the backbone of this dissertation.

Thank you to the scientific societies, who were very supportive of my research and career in many ways. Special thanks the arachnology societies and their meetings: the International Society of Arachnology (and the Oscar and Jan Francke fund), the American Society of Arachnology, and the Latin American Congresses of Arachnology. Thank you to the Pan American Society for Evolutionary Developmental Biology, the Society for Developmental Biology and The Society for Integrative and Comparative Biology. Thank you, the University of Wisconsin-Madison, the Department of Integrative Biology and the National Science Foundation for financially supporting my doctoral studies.

I could not have done anything of this and passed through the pandemic without the help of my family and friends. My immense gratitude to all the past and present housemates in Hypatia coop, where I lived for five years. Thank you to all my longtime and new friends in Madison, including Tiago, Henrique, Bruno, Julia, Page, Celso, Tyler, Don, Isac, all the Volleyball crew, Amelie, Zakher, Quinn and Sagara. My many thanks to all my longtime friends in Brazil. Even if just briefly seeing you during my summer visits, this always had a refreshing effect in me! Thank you!

My mom, Ana Paula, and dad, Osvaldo, have been amazing parents and none of this would have been possible without them. Growing older and becoming friends with your parents is a real privilege. My parents have always been very supportive of my education and career decisions, and I cannot think of a single action by my parents that did not push me further in keeping my studies and following a career as a biologist. Thank you for being my emotional

support in difficult times and helping me see the bright side of things. A career in science is unfortunately a very privileged path, and I can only be here because of the sacrifices and investments they made, from providing me access to the best schools throughout my life, to same-day flights with me to Rio to get a last-minute US visa interview. I know, six years was a long time. I want to make sure we spend always more time together and share many more moments. Thanks to my brother, Gabriel, for always being an uplifting person and spending time with me when I am in Brazil! I missed you a lot and I hope we can spend more time together as well! I love you all!

Lastly, I have so much to thank my partner, Tatiana. I still can't believe I could convince her to join me so far from home moving from Brazil. I am the luckiest person on Earth! Tati has been with me through all my ups and downs during this journey, and the most important reason why I was able to stay motivated and happy being so far from home and my friends, my culture, and the warmth of Brazil. In these years, I learned a lot more about myself and life because I am spending my time with her. Tati, I am so proud of everything you have achieved in your work in conservation and science, and you are a great inspiration to me every day. No science here without Tati. Thank you so much, for being with me, laughing about my bad jokes, and being so special. Te amo!

You will notice I, as a formality, have used the first person throughout the following introduction while briefing the findings of this dissertation. However, as seen in the list of authors in each chapter, and the acknowledged equal contribution notes, all the ensuing findings of this dissertation are the collective effort of the Sharma lab members and numerous collaborators in other institutions.

Dissertation Abstract

The sensory biology of chelicerates has been intensively studied with respect to morphology, physiology, and behavioral ecology. Nonetheless, relatively little is known about the development and macroevolution of chelicerate sensory systems, with emphasis on the genetic mechanisms underpinning the evolution of appendages specialized in sensing the world (e.g., pedipalps and antenniform legs), and the diversity of eye forms across arachnids. In this dissertation, I investigate body plan, appendage, and eye development in chelicerates. In Chapter 1, I establish a modern embryonic staging system for the emerging evodevo model, the daddy-longlegs *Phalangium opilio*. In Chapters 2 and 3, I present the first genome assembly for *P. opilio*, and elucidate the role of Hox genes in establishing the identities of chelicerae, pedipalps and legs in *P. opilio*, providing a foundation to study further appendage specializations in the group. In Chapter 2, I also provide evidence that the convergent evolution of supernumerary tarsomeres (subdivisions of the tarsus) in arachnids and insects relies on the co-option of Epidermal Growth Factor Receptor signaling. In Chapter 4, I establish genomic resources and protocols to study the development of antenniform legs in the whip spider *Phrynus marginemaculatus* and show that whip spiders retain signal of whole-genome duplication shared with spiders and scorpions. In Chapter 5, I investigate the genetic basis of eye reduction in troglobitic cave whip spiders and provide the first functional evidence that a Retinal Determination Network (RDN) gene is necessary for eye development in an arachnid, the spider *Parasteatoda tepidariorum*. Lastly, in Chapter 6, I report the discovery of rudimentary eye pairs in extant daddy-longlegs, a group thought to have only a single pair of median eyes. This discovery has implications for the phylogenetic placement of four-eyed fossil harvestmen, the sensory biology of Opiliones, and the evolution of eyes in Chelicerata.

Introduction

Chelicerates are one of the megadiverse radiations of Arthropoda and constitute fascinating animals that have existed on Earth's waters and land since the Cambrian (Edgecombe, 2020; Giribet and Edgecombe, 2019). The traditionally recognized extant lineages of this group include the sea spiders (Pygnogonida), the horseshoe crabs (Xiphosura), and an assemblage of terrestrial orders known as Arachnida, though recent studies suggest that horseshoe crabs are part of the arachnid radiation (Ballesteros and Sharma, 2019; Ballesteros et al., 2022). Arachnida is by far the most species-rich division of chelicerates, largely because of the more than 50,000 species of spiders (order Araneae) (World Spider Catalog, 2023) and 55,000 species of mites and ticks (various groups in the orders Acariformes and Parasitiformes) (Zhang, 2011). Nonetheless, a swath of shapes, behaviors and unique adaptations are also represented in nine smaller arachnid orders. These are the harvestmen (Opiliones; also known as daddy-longlegs), scorpions (Scorpiones), pseudoscorpions (Pseudoscorpiones), vinegaroons (Thelyphonida), short-tailed whip scorpions (Schizomida), whip spiders (Amblypigi), hooded tick-spiders (Ricinulei), palpigrades (Palpigradi), and camel spiders (Solifugae) (Foelix, 2011; Harvey, 2007; Pinto-da-Rocha et al., 2007; Punzo, 1998; Savory, 1964; Weygoldt, 2000). An association between humans and arachnids has been recorded since antiquity. For instance, references to scorpions that feature their distinct shape, fearsomeness, and potent venom are present in ancient Greek and Egyptian mythology, and these survive today as the name of a constellation (Melic, 2002). Spiders, with their dexterous eight legs and web-weaving behavior, are a memorable cultural staple from ancient Greece to contemporaneous superhero pop culture (Da-Silva et al., 2014; Melic, 2002). Moreover, various chelicerate taxa also have medical, agricultural, and economic significance in society. For instance, numerous mite and tick species are pathogen vectors or crop pests. Horseshoe crab blood is the source for an important clinical test (Limulus Amebocyte Lysate assay) essential for the pharmaceutical

industry, including a role in the development of COVID19 vaccines, during the pandemic that occurred in the timespan of this Ph.D. research (Arnoud, 2020; Walls et al., 2002).

Arguably one of the most fascinating aspects of chelicerates is their sensory biology. Multiple eye pairs, numerous appendages, and an array of sensory "hairs" (more aptly termed setae) set their perception of the world as something alien to humans. The eyes of some jumping spiders (Salticidae) have one of the highest visual acuities of the animal kingdom, with different pairs of eyes subdividing visual functions (Harland et al., 2012; Land, 1985). Foraging, mating, and communication in arachnids relies on a complex mix of mechanical, chemical and visual stimuli (Hebets, 2005; Hebets et al., 2013; Santer and Hebets, 2008). Spider mechanoreceptors, particularly in the wandering spider *Cupiennius salei*, have served as powerful models for studying the physical limits of vibration and airflow detection in animals, which are achieved by highly sensitive lyriform organs (grouped slit sensilla) and trichobothria (thin setae) (Barth, 2000; Barth, 2002; Barth, 2004; Foelix, 1985). Works on the visual system of the horseshoe crab *Limulus polyphemus* pioneered much of our understanding about the physiology and molecular basis of vision in arthropods (Battelle, 2017; Smith et al., 1995). While the sensory biology of arachnids has been thoroughly investigated with respect to the morphology of sensory organs, physiology and behavior, a fundamental dimension of their sensory biology which has received considerably less attention is the development and macroevolution of their sensory systems.

The sensory systems of chelicerates are intimately associated with their body plan. In other arthropods (insects, crustaceans, myriapods) most sensory structures involved in sensing chemicals, touch, and changes in temperature and humidity are generally concentrated in one or two pairs of head appendages, the antennae (Altner and Prillinger, 1980; Steinbrecht, 1984). The chelicerate body segments and appendages, on the other hand, are "partitioned" very differently. First, with respect to an insect, the positional homolog of the antenna in chelicerates

is the chelicera (Telford and Thomas, 1998), the namesake of this group, which is primarily used for feeding. In arachnids, these are followed by a pair of pedipalps, versatile appendages that have assumed an array of functions in different orders, from prey capture (e.g., whip spiders, vinegaroons) (Shultz and Pinto-da-Rocha, 2007; Weygoldt, 2000), to sperm transfer (Talarico et al., 2008; Foelix, 2011) and sensing the world (Willemart et al., 2009). Following the pedipalps, there are four pairs of walking legs. It has been argued that despite lacking antennae, arachnids tend to concentrate most of their sensory setae on the extremities of the appendages, particularly the pedipalps and the first and second pairs of legs (Foelix, 1985). This tendency is manifested in the convergent evolution of "antenniform legs" in some orders; for example, the first leg pair of whip spiders and the second leg pair of daddy-longlegs (particularly in Phalangida) are elongated, bear additional pseudo-segments (tarsomeres), and concentrate sensory setae of multiple modalities (Foelix and Hebets, 2001; Gainett et al., 2017; Santer and Hebets, 2011; Willemart et al., 2009). This phenomenon, in turn, is a striking convergence with the antenna of non-chelicerate arthropods. Understanding how sensory appendages evolved required understanding how the basic types of appendages (chelicera, pedipalp, and legs) are specified in the first place.

The past two decades have seen a resurgence of interest in the developmental biology of arachnids, with an Evolutionary Developmental Biology (evodevo) approach incorporating a comparative framework of molecular biology and genetics in these non-model organisms. Major insights into chelicerate body and appendage patterning have been achieved through embryological and genetic studies of spiders, and to a much lesser extent, of a mite model (Damen et al., 1998; Khadjeh et al., 2012; Pechmann et al., 2015; Schwager et al., 2017; Telford and Thomas, 1998). However, many questions concerning the evolution of novel sensory appendages (e.g., antenniform legs) can only be answered through study of groups with these dedicated sensory legs, such as daddy-longlegs and whip spiders. A further

confounding variable incurred by the predominance of spiders as models for understanding chelicerate body plan evolution is the occurrence of a genome duplication in this lineage (Schwager et al., 2017). This duplication was suggested to be absent in some arachnids, such as daddy-longlegs (Opiliones), and unduplicated genomes are understood to be the ancestral configuration of chelicerates (Schwager et al., 2017; Sharma et al., 2012a). The recently emerged model daddy-longlegs *Phalangium opilio*, the brain- and soul-child of my advisor Prashant Sharma (Sharma et al., 2012b; Sharma et al., 2012a; Sharma et al., 2013; Sharma et al., 2015), was therefore the most promising organism in the world to explore the evolution of the chelicerate body plan and sensory appendages. And so, I was very fortunate to be granted a chance to work at the University of Wisconsin-Madison in the Sharma lab, to explore these research topics and advance *P. opilio* as a research organism.

The first three chapters of this dissertation pertain to the topic of body plan evolution and appendage specification in chelicerates. In Chapter 1, I introduce a modern developmental staging system for the emerging evodevo model *P. opilio*. This study standardizes the terminology to study the embryology of this species and identifies some of the unanswered questions in Opiliones embryology. Chapter 2 provides an important milestone for the study of Opiliones and *P. opilio* as an emerging model for evodevo: the first genome assembly and annotation for this group of arachnids. This chapter also provides unambiguous evidence for an unduplicated genome condition in daddy-longlegs and provides the first functional data for Hox genes in an arachnid with an unduplicated genome. I show that the Hox gene *Deformed* (*Hox4*) and *Sex combs reduced* (*Hox5*) have canonical homeotic functions and are necessary for conferring leg identity embryonically. Leveraging the new genomic resources, I also investigate the genetic patterning of the supernumerary leg tarsomeres, which are most numerous in the sensory legs. My results suggest that that co-option of Epidermal Growth

Factor Receptor signaling underlies the convergent evolution of supernumerary tarsomeres (distal leg subdivisions) in daddy-longlegs and insects.

Chapter 3 focuses on investigating the genetic patterning of the chelicera and pedipalp. I provide evidence that the Hox gene *labial* (*Hox1*) has a dual function in segment specification and pedipalp identity, constituting the first functional evidence in Arthropoda that *labial* is an embryonic homeotic gene. The combined functional data of Chapters 2 and 3, in tandem with comparative surveys of Hox paralogs in embryos of a scorpion and a horseshoe crab (Chapter 3), enables the proposal of a functionally validated model of chelicerate anterior appendage specification.

In Chapter 4, I establish the foundations for investigating the evolution of antenniform legs in whip spiders (Amblypygi). Apart from a landmark study on the embryology of the Floridian whip spider *Phrynus marginemaculatus* almost 50 years ago (Weygoldt, 1975), there had been no studies on the embryology of Amblypygi, and virtually no information on their genomics and genetics. I generate genomic resources and tools to study the development and genomics of *P. marginamaculatus*, which enable the first gene expression and genomic surveys in an amblypygid. I show that whip spiders retain a signature of a genome duplication shared with spiders (Araneae) and have duplicated Hox genes. In addition, an expression survey of selected insect antennal genes (leg gap genes) suggests that the mechanism of leg antennification in this group is not convergent with the insect antenna.

Chapters 5 and 6 focus on a different sensory system: eyes. My focus here was to investigate both the genetic underpinnings of chelicerate eye formation, as well as the reduction of eyes as seen in cave-dwelling arachnids (Aharon et al., 2019; Mammola and Isaia, 2017) and in the arachnid order Opiliones. Chapter 5 provides a transcriptomic comparison of conserved eye-patterning genes in a sister-species pair of eye-bearing and eyeless Israeli cave whip spiders, which yielded targets for functional investigation of cave blindness in arachnids.

I also provide the first functional evidence that a Retinal Determination Network (RDN) gene, a paralog of the homeobox gene *sine oculis* (*Six-1*), is necessary for eye patterning in spiders.

Lastly, Chapter 6 begun as an exploration of eye patterning by RDN genes in the daddy-longlegs *P. opilio*, both because of its unduplicated genome conduction as a point of comparison to previous data in spiders (Samadi et al., 2015; Schomburg et al., 2015), and to search for clues of the mechanism behind the loss of lateral eyes in this lineage (spiders have both median and lateral eyes). Chapter 6 feels very special to me. It threads a line from the beginning of my career as an undergraduate researcher in Brazil with my former advisor, Rodrigo Willemart (when I was studying the anatomy of sensory setae and systematics of Neotropical harvestmen), into the comparative evodevo framework I trained on with my current advisor, Prashant Sharma, and beyond to potential novel directions on the evolution of eyes in Opiliones and Chelicerata. In this chapter, I report the serendipitous discovery of rudimentary eye pairs in daddy-longlegs, which I argue has important implications for the placement of fossils, sensory biology, and the evolution of eyes in Opiliones and Chelicerata.

This dissertation shows there is still much to learn about the evolution of development and sensory biology of chelicerates, and that the study of non-model organisms is exciting and bound to provide fresh insights into the study of animal evolution.

References

Aharon, S., Ballesteros, J. A., Crawford, A. R., Friske, K., Gainett, G., Langford, B., López, C. E. S., Ya'aran, S., Gavish-Regev, E. and Sharma, P. P. (2019). The anatomy of an unstable node: a Levantine relict precipitates phylogenomic dissolution of higher-level relationships of the armoured harvestmen (Arachnida: Opiliones: Laniatores). *Invert. Systematics* 33, 697–717.

Altner, H. and Prillinger, L. (1980). Ultrastructure of Invertebrate Chemo-, Thermo-, and Hygroreceptors and Its Functional Significance. In *International Review of Cytology* (ed. Bourne, G. H. and Danielli, J. F.), pp. 69–139. Academic Press.

Arnoud, C. (2020). Horseshoe crab blood is key to making a COVID-19 vaccine—but the ecosystem may suffer. *National Geographic*.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440 –22.

Ballesteros, J. A., Santibáñez-López, C. E., Baker, C. M., Benavides, L. R., Cunha, T. J., Gainett, G., Ontano, A. Z., Setton, E. V. W., Arango, C. P., Gavish-Regev, E., et al. (2022). Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Mol. Biol. Evol.* 39, msac021.

Barth, F. G. (2000). How to catch the wind: spider hairs specialized for sensing the movement of air. *Naturwissenschaften* 87, 51–58.

Barth, F. G. (2002). Spider senses–technical perfection and biology. *Zoology* 105, 271–285.

Barth, F. G. (2004). Spider mechanoreceptors. *Curr. Opin. Neurobiol.* 14, 415–422.

Battelle, B.-A. (2017). Opsins and their expression patterns in the xiphosuran *Limulus polyphemus*. *Biol. Bull.* 233, 3–20.

Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. and Tautz, D. (1998). A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl. Acad. Sci. USA* 95, 10665–10670.

Da-Silva, E., Coelho, L., Campos, T. M. de, Carelli, A., Miranda, G. de, Santos, E. L. dos, Silva, T. R. and Passos, M. I. dos (2014). Marvel and DC Characters Inspired by Arachnids. *Comics Grid.* 4, 1–14.

Edgecombe, G. D. (2020). Arthropod Origins: Integrating Paleontological and Molecular Evidence. *Annu. Rev. Ecol. Evol. Syst.* 51, annurev-ecolsys-011720-124437-25.

Foelix, R. F. (1985). Mechano- and chemoreceptive sensilla. In *Neurobiology of Arachnids*, pp. 118–137. Springer-Verlag Berlin Heidelberg.

Foelix, R. (2011). *Biology of Spiders*. Third Edition. Oxford University Press.

Foelix, R. and Hebets, E. (2001). Sensory biology of whip spiders (Arachnida, Amblypygi). *andrias* 15, 129–140.

Gainett, G., Michalik, P., Müller, C. H. G., Giribet, G., Talarico, G. and Willemart, R. H. (2017). Ultrastructure of chemoreceptive tarsal sensilla in an armored harvestman and evidence of olfaction across Laniatores (Arachnida, Opiliones). *Arthropod Struct. Dev.* 46, 178–195.

Giribet, G. and Edgecombe, G. D. (2019). The phylogeny and evolutionary history of arthropods. *Curr. Biol.* 29, R592–R602.

Harland, D. P., Li, D. and Jackson, R. R. (2012). How jumping spiders see the world. In *How Animals See the World: Comparative Behavior, Biology, and Evolution of Vision* (ed. Lazareva, O. F., Shimizu, T., and Wasserman, E. A.), pp. 132–163. Oxford Academic.

Harvey, M. S. (2007). The smaller arachnid orders: Diversity, descriptions and distributions from Linnaeus to the present (1758 to 2007). *Zootaxa* 1668, 363–380.

Hebets, E. A. (2005). Attention-altering signal interactions in the multimodal courtship display of the wolf spider *Schizocosa uetzi*. *Behav. Ecol.* 16, 75–82.

Hebets, E. A., Gering, E. J., Bingman, V. P. and Wiegmann, D. D. (2013). Nocturnal homing in the tropical amblypygid *Phrynus pseudoparvulus* (Class Arachnida, Order Amblypygi). *Anim. Cogn.* 17, 1013–1018.

Khadjeh, S., Turetzek, N., Pechmann, M., Schwager, E. E., Wimmer, E. A., Damen, W. G. M. and Prpic, N. M. (2012). Divergent role of the Hox gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl. Acad. Sci. USA* 109, 4921–4926.

Land, M. F. (1985). The Morphology and Optics of Spider Eyes. In *Neurobiology of Arachnids*, pp. 53–78. Springer-Verlag Berlin Heidelberg.

Mammola, S. and Isaia, M. (2017). Spiders in caves. *Proc. R. Soc. B* 284, 20170193–10.

Melic, A. (2002). De madre araña a demonio escorpión: Arácnidos en la mitología. *Rev. Iber. Aracnol.* 5, 112–124.

Pechmann, M., Schwager, E. E., Turetzek, N. and Prpic, N.-M. (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc. R. Soc. B* 282, 20151162.

Pinto-da-Rocha, R., Machado, G. and Giribet, G. (2007). *Harvestmen: The Biology of Opiliones*. Harvard University Press.

Punzo, F. (1998). *The Biology of Camel-Spiders*. Springer US.

Samadi, L., Schmid, A. and Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *EvoDevo* 6, 16.

Santer, R. D. and Hebets, E. A. (2008). Agonistic signals received by an arthropod filiform hair allude to the prevalence of near-field sound communication. *Proc. R. Soc. B* 275, 363–368.

Santer, R. D. and Hebets, E. A. (2011). The Sensory and Behavioural Biology of Whip Spiders (Arachnida, Amblypygi). In *Advances in Insect Physiology* (ed. Casas, J.), pp. 1–64. Burlington: Academic Press.

Savory, T. (1964). *Arachnida*. Academic Press Inc.

Schomburg, C., Turetzek, N., Schacht, M. I., Schneider, J., Kirfel, P., Prpic, N.-M. and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* 6, 15.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012a). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012b). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* 14, 522–533.

Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol. Dev.* 15, 228–242.

Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C. and Extavour, C. G. (2015). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. B* 282, 20150698.

Shultz, J. W. and Pinto-da-Rocha, R. (2007). Morphology and functional anatomy. In *Harvestmen: The Biology of Opiliones* (ed. Pinto-da-Rocha, R., Machado, G., and Giribet, G.). Harvard University Press.

Smith, W. C., Greenberg, R. M., Calman, B. G., Hendrix, M. M., Hutchinson, L., Donoso, L. A. and Battelle, B. (1995). Isolation and Expression of an Arrestin cDNA from the Horseshoe Crab Lateral Eye. *J. Neurochem.* 64, 1–13.

Steinbrecht, R. A. (1984). Chemo-, Hygro-, and Thermoreceptors. In *Biology of the Integument*, pp. 523–553. Springer, Berlin, Heidelberg.

Talarico, G., García Hernández, L. F., Michalik, P. The male genital system of the New World Ricinulei (Arachnida): Ultrastructure of spermatozoa and spermiogenesis with special emphasis on its phylogenetic implications. *Arthropod Struct. Dev.* 37, 396–409.

Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl. Acad. Sci. USA* 95, 10671–10675.

Walls, E. A., Berkson, J. and Smith, S. A. (2002). The horseshoe crab, *Limulus polyphemus*: 200 million years of existence, 100 years of study. *Rev. Fish. Sci.* 10, 39–73.

Weygoldt, P. (1975). Untersuchungen zur Embryologie und Morphologie der Geißelspinne *Tarantula marginemaculata* C. L. Koch (Arachnida, Amblypygi, Tarantulidae). *Zoomorphologie* 82, 137–199.

Weygoldt, P. (2000). *Whip Spiders (Chelicerata: Amblypygi). Their Biology, Morphology and Systematics*. Apollo Books.

Willemart, R. H., Farine, J.-P. and Gnaspini, P. (2009). Sensory biology of Phalangida harvestmen (Arachnida, Opiliones): a review, with new morphological data on 18 species. *Acta Zool.* 90, 209–227.

World Spider Catalog (2023). World Spider Catalog. Version 24. Natural History Museum Bern, online at http://wsc.nmbe.ch, accessed in April 2023. doi: 10.24436/2

Zhang, Z.-Q. (2011). Animal biodiversity: An introduction to higher-level classification and taxonomic richness. *Zootaxa* 3148, 7–12.

Chapter 1

Eggs to long-legs: embryonic staging of the harvestman *Phalangium opilio* (Opiliones), an emerging model arachnid

Guilherme Gainett†, Audrey R. Crawford†, Benjamin C. Klementz, Calvin So, Caitlin M. Baker, Emily V. W. Setton and Prashant P. Sharma*

*Correspondence: guilherme.gainett@wisc.edu
†Guilherme Gainett and Audrey R. Crawford contributed equally to this work
Department of Integrative Biology, University of Wisconsin-Madison, 438 Birge Hall, 430 Lincoln Drive, Madison, WI 53706, USA

Published March 2022 in Frontiers in Zoology

Volume 19, Article number: 11, doi.org/10.1186/s12983-022-00454-z

Abstract

Background: The comparative embryology of Chelicerata has greatly advanced in recent years with the integration of classical studies and genetics, prominently spearheaded by developmental genetic works in spiders. Nonetheless, the understanding of the evolution of development and polarization of embryological characters in Chelicerata is presently limited, as few non-spider species have been well studied. A promising focal species for chelicerate evo-devo is the daddy-long-legs (harvestman) *Phalangium opilio*, a member of the order Opiliones. *Phalangium opilio*, breeds prolifically and is easily accessible in many parts of the world, as well as tractable in a laboratory setting. Resources for this species include developmental transcriptomes, a draft genome, and protocols for RNA interference, but a modern staging system is critically missing for this emerging model system.

Results: We present a staging system of *P. opilio* embryogenesis that spans the most important

morphogenetic events with respect to segment formation, appendage elongation and head development. Using time-lapse imaging, confocal microscopy, colorimetric in situ hybridization, and immunohistochemistry, we tracked the development of synchronous clutches from egg laying to adulthood. We describe key events in segmentation, myogenesis, neurogenesis, and germ cell formation.

Conclusion: Considering the phylogenetic position of Opiliones and the unduplicated condition of its genome (in contrast to groups like spiders and scorpions), this species is poised to serve as a linchpin for comparative studies in arthropod development and genome evolution. The staging system presented herein provides a valuable reference for *P. opilio* that we anticipate being useful to the arthropod evo-devo community, with the goal of revitalizing research in the comparative development of non-spider arachnids.

Keywords: Development, Chelicerata, Evo-devo, *vasa, engrailed*, Eupnoi

Background

Comparative studies with rich taxonomic sampling are essential to understanding the evolution of development. For instance, in arthropods, the molecular mechanisms of simultaneous formation of segments in the fruit fly *Drosophila melanogaster* by means of gap, pair-rule, and segment polarity genes was initially established as a paradigm of arthropod segmentation (Rosenberg et al., 2009). Subsequent investigation of alternative model systems across different arthropod groups later supported the hypothesis that sequential addition of segments mediated by Notch signaling is the ancestral mode of segmentation in Arthropoda (Rosenberg et al., 2009; Stahi and Chipman, 2016). The discovery of a vertebrate-like oscillatory mechanism of posterior segment addition in an array of non-dipteran arthropods ignited fierce debate over the homology of segmentation in Bilateria (Chipman, 2010; Damen, 2007; Seaver EC., 2003; Stollewerk et al., 2003a). Key to the phylogenetic polarization of character states within the phylum Arthropoda, and specifically, at the root of the phylum, is Chelicerata, the sister group to the remaining arthropods.

Studies on the embryology of Chelicerata flourished in the first half of the twentieth century, with classical descriptive and experimental works that have laid the foundations for the evolution of development in this group (reviewed by [Anderson, 1973; Yoshikura, 1975]). Unification of chelicerate embryology with developmental genetics has, however, accelerated just in the last two decades (Abzhanov et al., 1999; Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006; Akiyama-Oda and Oda, 2010; Akiyama-Oda and Oda, 2020; Barnett and Thomas, 2012; Barnett and Thomas, 2013; Damen et al., 1998; Pechmann et al., 2015; Schwager et al., 2017; Sharma et al., 2014a; Telford and Thomas, 1998). These studies, mostly centered on two focal spider species, have led to a renaissance in the field of chelicerate development and revealed novel aspects of their evolutionary history. As an example, a modern reinvestigation of the classic graft experiment of axis specification in spiders (Holm, 1952)

identified a genetic mechanism shared with insect dorso-ventral axis patterning (i.e., *decapentaplegic* signaling) as the basis for the activity of the cumulus, an axis organizer in spiders (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006; Pechmann, 2020; Pechmann et al., 2017).

Nevertheless, the embryology of non-spider arachnids has received considerably less attention in the recent literature. Orders such Pseudoscorpiones, Thelyphonida (vinegaroons), and Amblypygi (whip-spiders) suffered from decades of hiatus in contributions to their embryology (but see [Ontano et al., 2021] and [Gainett and Sharma, 2020]). Embryogenesis in other orders, such as Ricinulei (hooded tick-spiders), Palpigradi (microwhip scorpions), and Solifugae (camel spiders), remains virtually unknown (Anderson, 1973; Schwager et al., 2015b). A modern understanding of the genetic mechanisms underlying the development of non-spider arachnids has great potential to elicit evidence for major evolutionary transitions within Chelicerata, including the parallel evolution of aerial respiratory structures and the specification of different types of appendages (Schwager et al., 2015b; Sharma, 2017).

One arachnid species that has emerged in the evo-devo literature as a focal taxon for comparative studies is the daddy-long-legs *Phalangium opilio* Linnaeus 1758 (Baudouin-Gonzalez et al., 2021; Gainett et al., 2021; Garwood et al., 2014; Leite et al., 2018; March et al., 2018; Setton and Sharma, 2018; Setton et al., 2017; Sharma, 2017; Sharma et al., 2012a; Sharma et al., 2012b; Sharma et al., 2013; Sharma et al., 2014b; Sharma et al., 2015a), a member of the order Opiliones (commonly known as “harvestmen”), and the first species of Opiliones described with the Linnean system. Opiliones is the third-largest order of Arachnida (sensu [(Ballesteros and Sharma, 2019)]; arachnids likely include horseshoe crabs), with over 6,500 described species (Giribet and Sharma, 2015; Kury et al., 2021). Among the remarkable features of this group are unique repugnatorial glands for deterring predators (Gnaspini and Hara, 2007), multiple origins of parental care, and highly diverse, sexually dimorphic armature

(Buzatto and Machado, 2014; Machado and Macías-Ordónéz, 2007), as well as an elongated pair of legs that are modified to serve a sensory function (Gainett et al., 2017a; Gainett et al., 2017b; Willemart et al., 2009). The embryology of a few Opiliones species was described in detail based on light microscopy and illustrations by studies published in the twentieth century (Holm, 1947; Juberthie, 1961; Juberthie, 1964; Moritz, 1957; Moritz, 1959; Muñoz-Cuevas, 1971) (for historical review, see [Gnaspini P, 2007]) The development of *P. opilio* was first characterized by Moritz (Moritz, 1957)but references to embryonic stages in recent studies on *P. opilio* development (e.g., [Sharma et al., 2012b, 2012a]) are based on the staging system provided by Juberthie (Juberthie, 1964) for the phalangiid species *Odiellus gallicus*[1]. Beyond this discrepancy in taxonomy, a description and staging of the development of any Opiliones species using modern approaches is still missing. Such description would aid in comparisons with the better-studied spider systems, namely *Parasteatoda tepidariorum* and *Cupiennius salei* (Mittmann and Wolff, 2012; Wolff and Hilbran, 2011), and serve as a resource for functional works on Opiliones developmental genetics.

Herein, we present a staging system of *P. opilio* embryogenesis using confocal microscopy, in situ hybridization, and time-lapse imaging. We tracked the development of synchronous clutches from egg laying to adulthood and described key events in neurogenesis, myogenesis, and segmentation of the body and appendages.

Results

We aimed to maximize compatibility between the stage numbers we propose here for *P. opilio* with the detailed description of Juberthie (1964) in *O. gallicus*[1], and particularly the descriptions of *P. opilio* embryos by Moritz (1957) and Winkler (1957). For defining stages, we combined both the characteristics from the external morphology that may be observed through the vitelline membrane, as well as specific features more clearly visible in dissected

embryos using fluorescent microscopy. We used both molecular and morphological markers to characterize the germ cells and describe their dynamics. We optimized antibody staining protocols for *P. opilio* and complemented the description of some of the reference stages with cross-reactive antibody staining for muscle cells (Tm1; tropomyosin) and neural cells (acetylated α-tubulin).

Overview of development

Females of *P. opilio* lay one egg clutch approximately every three days, when provided with fresh food and a container with moist coconut fiber at 26 °C (Fig. 1C–E; Additional file 1: File S1). Upon pairing, males and females usually readily mate after a few seconds (Fig. 1B). Adult animals can be collected in temperate regions of North America, Europe, East Asia, as well as New Zealand. In north America (data available from Massachusetts and Wisconsin), adults are found in early June and the breeding season ends around early September. Nonetheless, we can successfully maintain overwintering colonies in the laboratory, and new adult females are able to breed and lay eggs during winter months. The adult lifespan is 40–60 days.

A typical egg clutch contains approximately 250 embryos (Fig. 1D–E) that develop nearly synchronously. Eggs measure approximately 550 μm in diameter, are light-yellow in color, and are almost spherical. The egg has a thin, opaque chorion and a thick, transparent vitelline membrane. The chorion is easily removed with 100% commercial bleach for fixations (see Methods). Most embryos dechorionated with 50% bleach solution can continue development and hatch in halocarbon oil or 1 × phosphate buffered saline (PBS). Upon laying, embryos from this population require around 24 days (582 h, clutch #4) to hatch at 26 °C. A summary figure of all embryonic stages, diagnostic features and timing is provided in Additional file 2: File S2.

Embryonic staging

(a) Stages 1–3: Initial cleavages (days 1–4).

Stages encompassing the first cleavages and a blastoderm with loosely arranged yolk globules are here collectively treated as stages 1–3 until further study (see Discussion), following Fig. 20 of Juberthie (1964) in the description of *Odiellus gallicus*[1]. Embryos in this range proved challenging to investigate in *P. opilio*. In very early embryos, yolk cells and embryonic cells present a similar coloration and thus the initial cleavages were not observed at these early stages through the vitelline membrane. Attempts to manually remove the vitelline membrane in fixed embryos to observe cells at stages before the blastoderm formation resulted in the disintegration of the embryo, suggesting that the cells have limited cohesion at this stage, even after extended fixation periods. The broad range of autofluorescence of this membrane hampers visualization of cells under a fluorescent stereomicroscope with Hoechst and GFP. From egg laying to 72 h after egg laying (hAEL) (stages 1–3), we observed loosely arranged, large yolk globules at the periphery of the embryo (Fig. 2A). We also observed the overall dynamic of the cells in these early three stages by performing time-lapse imaging of 3-day-old embryos. Before the formation of the perivitelline space (see below), we observed movement of the yolk globules (Additional file 3: File S3).

(b) Stages 4–6: Late blastoderm and germ band (day 5; 120–147 hAEL)

Stage 4: Late blastoderm

A stage 4 embryo is diagnosed by the first sign of a perivitelline space. When approaching 120h (5 days), the embryo cells look more cohesive and the perivitelline space appears. (Fig. 2A; Additional file 3, 4: File S3– S4). Fixed embryos at this stage may be peeled

from the vitelline membrane and observed under a fluorescent stereomicroscope. A stage 4 embryo is nearly spherical and has one layer of cells with small nuclei surrounding the yolk globules (Fig. 3A, D). A few larger nuclei are present immediately below the outer layer of small nuclei and they are roughly regularly spaced (Fig. 3A– C). A denser elliptical spot of multilayered cells is observed (Fig. 3D, D′).

Stage 5: Early germ band (130 hAEL)

Stage 5 is defined by the formation of a horseshoe-shaped region on the blastoderm (Fig. 2B; Fig. 4A, A′; white arrow) and expansion of the perivitelline space (Fig. 2B; Additional file 3, 4: File S3–S4). The germ band is recognizable as a denser layer of cells on a hemisphere of the yolk, being the ventral side of the embryo, with an antero-posterior axis; the posterior end assumes a horseshoe shape (Fig. 4A, A′). At this stage, a whiter group of slightly elevated cells occurs at the posterior of the embryo (Additional file 4: File S4, frame 100h, yellow arrowhead). The anterior end has a denser layer of cells, at the place of the future cephalic region. No clear expression of the segment marker *Popi-engrailed* (*Popi-en*) could be detected by in situ hybridization at this stage (Fig. 4A, A′).

Stage 6: segmented germ band (140 hAEL)

At stage 6 the germ band is narrower than in stage 5 and has seven prosomal segments (i.e., the anterior tagma of chelicerates) and a posterior growth zone (Fig. 4B– E′). Stage 6 is therefore marked by the segmentation of the prosoma. The ocular segment (the most anterior) expresses *Popi-en* as a dot (Fig. 4E; asterisk). The other six segments express *Popi-en* as a stripe at their posterior edge (a conserved expression domain in Arthropoda (Fig. 4B–E′). Early stage 6 is nearly spherical and the *Popi-en* stripes of expression are thin and uniform across the left and right edges of the germ band (Fig. 4B–D′). The horseshoe-shaped posterior germ band

is sharply defined as a segment growth zone (Fig. 4C, C′; white arrow). The whiter spot of cells is surrounded by the horseshoe-shaped growth zone, posterior to it (Fig. 5A). These cells express the germ cell marker *Popi-vasa* throughout embryogenesis (Fig. 5A–D).

Primordia of the head lobes are formed, initially appearing as two dense spots (Fig. 4D, D′). At a late stage 6, the germ band extends posteriorly as each segment becomes broader, and the embryo assumes a more oval shape. Each prosomal segment becomes denser where the limb buds will protrude (Fig. 4E′; black arrows). The head lobes expand laterally at the anterior end of the germ band (Fig. 4E; dotted line). The stage numbers from this point up to stage 12 match the number of *engrailed* stripes posterior to the ocular segment. The convention of stages using segments as landmarks is particularly useful for studying *P. opilio* embryogenesis, because stages 6–13 encompass appendage formation and elongation, segmentation of the body, and head elaboration.

(c) Stages 7–14: elaboration of the head and formation of the opisthosoma (days 6–10; 147–258 hAEL)

Stage 7: first opisthosomal segment (147 hAEL)

Stage 7 is defined by the occurrence of the first opisthosomal segment, and the oblong shape of the embryo (Fig. 6). From this stage onwards, each of the nine opisthosomal segments form sequentially from the anterior border of the posterior growth zone. Each new segment expresses *Popi-en* at its posterior border (Fig. 6A–B). Stage 7 also encompass the initial formation of the prosomal limb buds (chelicera, pedipalp, L1-L4 legs), which slightly bulge from the germ band and are wider than longer (Fig. 6A, B, E). The expression of *Popi-en* concentrates on the lateral edges of each prosomal segment and appears fainter at the ventral midline of the embryo (Fig. 6A–B).

Stage 8: second opisthosomal segment, limb buds (153 hAEL)

Stage 8 has two opisthosomal segments formed and limb bud primordia with a proximo-distal axis (longer than wider) (Fig. 2D; Fig. 6C–D', F–F'). The limb bud of the second pair of legs is slightly longer than the other appendages (Fig. 6C–D', F–F') and this allometry is maintained from this stage to adulthood, when this elongated leg assumes a sensory function. The primordium of the stomodeum begins to form medially on the head lobes (Fig. 6F). The germband is narrower than in the previous stage (compare Fig. 6B and D). Each limb bud is composed of an external layer of 3–5 cells that do not stain for tropomyosin, and an internal layer of cells staining for tropomyosin (Fig. 6F′, F″). A space exists between these two layers of cells, and the growing limb buds maintain this space along the appendages as they elongate, so that the bundle of cells stained for tropomyosin traverses the tube-like appendage (Fig. 6F′, F″).

Stage 9: third opisthosomal segment (171 hAEL)

Stage 9 is defined by the presence of three opisthosomal segments (Fig. 7A–C). The horizontal stripes of *Popi-en* are interrupted along the developed ventral sulcus (Fig. 7B, B′). The labrum has begun to form but does not yet project posteriorly to cover the stomodeum. Cells expressing tropomyosin are present on the head lobes and on the developing labrum anterior to the stomodeum, which now is open (Fig. 7C). Cells expressing tropomyosin also occur as two bilaterally symmetrical horizontal bands at each body segment, below the outer layer of ectodermal cells and interrupted at the ventral sulcus (Fig. 7C). As the development of legs progresses, tropomyosin-positive cells form a continuous bundle from the base to the tip of each appendage. At late stage 8 (Fig. 6; black arrows) and more clearly from stage 9 onwards, neural precursor cells differentiate and invaginate in the ventral ectoderm and head

lobes. The apical process of invaginating cells appears as a bright spot of alpha-tubulin expression, which is enriched in neural cells (Fig. 7D–D′; black arrows). The sites of invagination are regularly spaced in the ventral ectoderm and are surrounded by 6–8 ectodermal cells, forming rosettes (Fig. 7D–D′; black arrows).

We further analyzed the germ cells at stage 9. The germ cells are distinctively larger than surrounding cells (Fig. 5E) and occur below the ectoderm (Fig. 5E). The nuclei of the germ cells have loosely compacted chromatin compared to the ectodermal cells (Fig. 5E′). The germ cell cluster at stage 9 occurs anterior to the growth zone and posterior to O3 (in contrast to posterior to the growth zone in stage 6). By the end of stage 9, the germ cell cluster is localized in the nascent O4 (Fig. 5B, E). From stage 9, until at least stage 14, the germ cell cluster remains within O4 and O5, as evidenced by the co-visualization of *Popi-en* and *Popi-vasa* (Fig. 5).

Stage 10: fourth opisthosomal segment (174 hAEL)

Stage 10 has four opisthosomal segments formed. The embryo is no longer C-shaped in lateral view and the opisthosoma begins to project ventrally (Fig. 2F; Fig. 8A– C). This ventral flexure may be observed in the time-lapse of developing embryos (Fig. 2F; Additional file 3: File S3). In ventral view, the yolk is visible only in the anterior half of the embryo (Fig. 8B). The bundle of tropomyosin-positive cells of the chelicera begins to branch, and externally the distal part of the appendage has a triangular shape, or a slight indentation, at late stage 10 (Fig. 8C–C′). A line of 10 cells positive for tropomyosin expression occur along the ventral sulcus spanning the prosomal segments 3–6 (Fig. 8C; white arrows). A few cells already appear in stage 9 (Fig. 7C; white arrows). We did not follow the fate of these cells further in development. The anterior and ectal rims of the head lobes begin to form the anterior furrow (cerebral fold sensu [Holm, 1947]; semilunar grooves, sensu [Anderson, 1973]), and a lateral

furrow (Fig. 8C; see also stage 11). The labrum resembles a “nose”, pointing posteriorly and covering the mouth in frontal view (Fig. 8A–C).

Stage 11: fifth opisthosomal segment (177 hAEL)

This stage is marked by the formation of the fifth opisthosomal segment. At early stage 11, L2 legs overlap crossing the ventral midline of the embryo and L3 and L4 legs are curved up (Fig. 9A–B′; 9F′). The distal podomere of the chelicera is branched (Fig. 9E–F). As in stage 10, the yolk is visible only in the anterior half of the embryo (Fig. 9B). The anterior margin of the head ectoderm moves towards the base of the chelicerae, folding over the neurogenic ectoderm of the head (Fig. 9E–G; asterisk) (completed by stage 13). At early stage 11, two slit invaginations occur adjacent to a medial bridge connecting the labrum and the anterior rim of the head (Fig. 9E). The anterior and lateral furrows deepen to surround the left and right lobes of the head, and a narrow medial bridge connects the labrum and the anterior fold (Fig. 9E). At late stage 11 the posterior margin of the folding tissue is orthogonal to the medial bridge, and the slit invaginations are still visible as two pits adjacent to the medial bridge (Fig. 9F).

Stage 12: sixth opisthosomal segment (216 hAEL)

Stage 12 is defined by the sixth opisthosomal segment. The anterior and lateral margins of the anterior fold cover the medial bridge and the paired slits, the latter not visible anymore in frontal view (Fig. 9C–D′, G–G′). The lateral margins of each segment expand dorsally around the yolk, so that in ventral view the dorsal yolk is not visible (Fig. 9D–D′). L2 legs cross over the germ band lateral margin on the opposite side of the embryo and curl anteriorly, while L4 legs cross the ventral midline and reach the coxae of L2 legs (Fig. 9D, D′, F). The proximal article of the chelicerae and the endites of the pedipalps do not overlap the labrum (Fig. 9D, D′, G′). The gonopores first appear at this stage flanking the ventral midline on the

second opisthosomal segment (Fig. 7C, D of [Sharma et al., 2012b]), whereas the spiracles open more laterally on the same segment (along the same longitudinal axis as the coxa-trochanter joints of the walking legs; Fig. 8A of [Sharma et al., 2012b]).

Stage 13: seventh opisthosomal segment (228 hAEL)

Stage 13 (Fig. 10) is defined by the seventh opisthosomal segment. The anterior fold of the head has completely covered the medial bridge (Fig. 10A–B′) and forms the prosomal margin (*clypeus* sensu [Holm, 1947]). The proximal article of the chelicera and the endites of the pedipalps overlap with the labrum (respectively, in frontal and ventral view) (Fig. 10B–C′). Late stage 13 has parallel chelicerae and the labrum ventral to them (Fig. 10C–C′).

Stage 14: final opisthosomal segments (246 hAEL).

Stage 14 (Fig. 11) is defined by the presence of the *corona analis*, a projection of the opisthosomal end comprising the opisthosomal segments 8–9 and the telson around the proctodeum (Fig. 11A) (see also [Winkler, 1957]). The chelicera has fixed and movable digits bearing pointed termini (Fig. 11A). The coxapophyses of the pedipalp and L1 legs meet at the ventral midline, forming the floor of the stomotheca (preoral chamber) Fig. 11A; dotted lines). The nervous system at this stage is well developed, with marked circumpharyngeal connectives to the brain and ventral nerve cord with parallel commissures (Fig. 10B) (not analyzed at previous stages).

(d) Stages 15–19: elaboration of the eyes, organogenesis, and cuticular characters (days 10–22)

According to Holm (1947) and Juberthie (1964), organogenesis begins with the end of segmentation, and proceeds until the embryo hatches. Formation of the digestive caeca and the

heart have been addressed in previous works (see [Juberthie, 1964; Moritz, 1957; Winkler, 1957]), and occur during stages 15–19 (see below) (Juberthie, 1964; Moritz, 1957; Winkler, 1957). We defined the final stages based on external landmarks visible in bright field, and that are evenly spaced in time from stage 14 up until hatching.

Stage 15: Eye pigmentation (day 10)

Stage 15 is defined by the onset of red pigment deposition in the eyes (Fig. 12B–C; black arrowheads). Pigment is initially restricted to the lower half of the eye (frontal view) and is light-red colored in bright field (Fig. 12B–C).

Stage 16: Egg tooth pigmentation (day 12)

Stage 16 is defined by the darkening of the tarsal claws in the pedipalp and L1–L4 legs (Fig. 12D–E; black arrows), and the darkening of the egg tooth, an embryonic structure anterior to the eyes in the margin of the prosoma (Fig. 12D–E; white arrow).

Stage 17: Ozopore pigmentation (day 14)

Stage 17 is defined by the onset of red pigmentation around the ozopores, which are the openings of the repugnatorial glands (Fig. 12F–H; white double-arrows). It is unclear if the pigmentation derives from the gland itself, or whether it is a pigmentation of the cuticle.

Stage 18: Lens formation (day 17)

Stage 18 is defined by the appearance of a distinct cuticular lens in the eye (Fig. 12I–J; white arrowheads). The lens is easily discernible because the pigmentation in the eye is black in the lens area, and grey-blue colored around it (Fig. 12I–J; compare to Fig. 12E–H). Patches of pigmentation in the dorsal prosomal cuticle start to become visible at this stage and darken

up until hatching (Fig. 12I–L).

Stage 19: Setae and chelicera claw darkening (day 19)

This last stage before hatching is marked by the darkening of setae on the appendages, and of the movable and fixed finger of the chelicerae (Fig. 12K–L; black double-arrows). Movements of the body and appendages are clearly visible.

(e) Hatchling and post-embryonic development

The term 'larva' has been used in the harvestman literature to describe the stage immediately after hatching (Gnaspini, 2007; Juberthie, 1964), because it lacks certain features of the next instars, such as setae. In the spider literature, the term was evaluated with some criticism by Wolff and Hilbrant (2011) due to the confusing homology schema with insects implied by its use. In their staging system of the ctenid spider *C. salei*, Wolff and Hilbrant (2011) proposed the alternative "postembryo" to describe the spider hatchling, which is immobile, lacks sensory hairs, and has only partially developed eyespots; only after molting from this stage does the spider become a first instar. This convention was adopted by subsequent investigations of spider development (e.g., [Mittmann and Wolff, 2012]). On the same grounds, we here adopt the term postembryo to refer to the hatchlings of *P. opilio*.

Embryos hatch into postembryos between 22 and 24 days after oviposition at 26ºC (Additional file 5: File S5). The postembryos of *P. opilio* have a flexible opisthosoma, which telescopes in and out periodically. In contrast to the spider postembryo, the hatchling of *P. opilio* is mobile, has well developed eyes, and has already produced setae under the postembryo cuticle. Within 2h of hatching, postembryos undergo the first molt and reach the first instar stage (Fig. 13A–C; Additional file 6: File S6). This fast molting is likely possible because of precocious formation of the first instar cuticle inside the postembryo cuticle while still inside

the egg (clearly visible in distal appendage termini; see Fig. 3 of [Gainett et al., 2021b]). The second instar is reached seven days after hatching (Fig. 13D). Six instars occur before the final molt into adulthood (Table 1), but this number may vary from 6–8 in different populations of this species (Gnaspini, 2007). Molting is approximately synchronous up to the 5th instar, when the first signs of sexual dimorphism in the male chelicera (horns) begin to appear. At the sixth instar, immature males have a more delicate body than immature females and exhibit primordia of the cheliceral horns. Of the two animals that survived to adulthood from the tracked clutch, one was a male and one a female. The male molted to adulthood at day 48 after hatching and the female at day 55 after hatching (Table 1).

Discussion

Early embryogenesis and germ cells

Only a few studies have documented the early cleavages and formation of the blastoderm in Opiliones, and many aspects remain unclear. According to Holm (1947), the first cleavages are believed to be intralecithal (Holm, 1947), which follows the general pattern of other arachnids (Anderson, 1973). Juberthie (1964) depicted a blastoderm stage composed of equally spaced sparse dots. These dots probably correspond to the ‘small white spots’ also described by Holm (1947), which are cells more deeply situated and larger than the peripheral blastoderm nuclei. Our observations of the late blastula (Stage 4) closely match that of Holm (1947) concerning the large subperipheral nuclei, which can be clearly observed with confocal microscopy.

Holm (1947) summarized the studies of Faussek (1891) and Schimkéwitsch (1898) on the early blastoderm stage: both authors described a blastoderm composed of peripheral cells and the differentiation of a mass of cells localized in a large white spot. Faussek (1891) interprets this large spot as the early genital rudiment, while Schimkéwitsch (1898) terms this

mass of cells the *cumulus primitivus*.

Schimkéwitsch's *cumulus primitivus* is composed of cells described to internalize into the yolk and of cells that will form the genital rudiment. A germ cell identity for this mass of cells was also proposed by Juberthie (1964) in the species *Odiellus galicus* (*massa genitale*), but he did not mention cell internalization at early stages. Adjacent to this initial mass of cells, Holm (1947) observed a second spot that separates from this first. He hypothesized that this second spot is the homologous structure to the true cumulus of spiders (Holm, 1947) (see below). According to Holm (1947) the genital rudiment becomes part of a larger horseshoe-shaped field where there is a groove that is likely the remnant of the blastopore. While these cells have not been reinvestigated until the present paper, the cumulus and germ cells of arachnids are best understood in spiders.

Spider embryos have a dense spot at the center of the germ disc, termed primary thickening (synonymous with *cumulus primitivus* and *cumulus anterior*). The primary thickening is the main site of cell internalization (gastrulation) (Akiyama-Oda and Oda, 2003; Mittmann and Wolff, 2012; Wolff and Hilbran, 2011) (but see Edgar et al., 2015). A subset of these internalized cells, termed cumulus (synonymous with *cumulus posterior* and cumulus mesenchymal cells) migrates under the ectoderm to the rim of the germ disc and establishes the dorso-ventral axis of the embryo (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006). The formation of the germ cells in the spider *P. tepidariorum* is unrelated to the cumulus: clusters of germ cells are only recognizable at the germ band stage and later as paired groups of cells in the opisthosomal segments expressing high levels of germ cell markers *vasa* and *piwi* (Schwager et al., 2015a).

Our observations of the daddy-long-legs *P. opilio* confirm the formation of a denser spot of cells on the late blastoderm (stage 4, see Fig. 3), which is likely the site of gastrulation (i.e., blastopore). Later, we also observed, by means of time-lapse imaging, a denser spot of

cells at the posterior end of the nascent germ band (Additional file 4: File S4). This mass of cells "moves" posteriorly, but this movement seems to be a consequence of the anteroposterior extension of the germband and coincides with the position where the putative primordial germ cells occur (see Juberthie (1964). While we cannot rule out the presence of a cumulus, we found no evidence of an independently moving cumulus at a stage prior to the polarized germ band (*contra* Holm, 1947). Furthermore, the germ cell marker *Popi-vasa* is expressed in the whiter cell cluster posterior of the horseshoe-shaped growth zone at stage 6 (Fig. 5), strongly suggesting their early identity as germ cells (also see below). The germ cell cluster is also characterized by larger cytoplasmic volume and loosely arranged chromatin relative to surrounding ectodermal and mesodermal cells (Fig. 5E–E′). We note that we may not yet rule out that the darker *Popi-vasa* staining at this cell cluster in early germ band stage is an artifact of the multi-layered nature of the cells, because in the spider *P. tepidariorum*, *vasa* is ubiquitously expressed in early germ disc embryos and causes the impression of higher expression at the multi-layered primitive plate (Schwager et al., 2015a).

Later in development, the germ cell cluster of *P. opilio* is positioned anterior to the growth zone (up until stage 9), and thereafter located on opisthosomal segment 4 (O4; stage 10). The observations suggest an anterior migration, as previously noted by Juberthie (1964) in *O. gallicus*[1]. In *P. opilio*, between stage 6 and stage 10, the germ cells migrate from the posterior to the anterior growth zone; next, they leave the growth zone to stay at the nascent O4, where they remain at least until stage 14, as also described by Moritz (1957).

Outstanding questions in the early embryogenesis of Opiliones include: (1) Where and how do the early cleavages occur? (2) What is the relationship between the initial mass of cells, the blastopore, and the germ cells? (3) Does a moving organizer (cumulus) occur in Opiliones? A detailed study of the genetic markers for the cumulus (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006; Akiyama-Oda and Oda, 2010; Pechmann et al., 2017) and germ

cells (Schwager et al., 2015a), in tandem with modern techniques for morphological investigation in *P. opilio,* offers an opportunity to provide definitive answers to these century-old questions. These future investigations will require overcoming the challenges of fixing early stages, or live imaging of the early stages with enhanced contrast and fluorescence markers through the eggshell.

Segmentation

The sequential segmentation of the prosoma was first documented in early works from the nineteenth century (reviewed in Holm, 1947), but a specific description of segment formation was only later addressed in *Phalangium opilio* (Moritz, 1957; Winkler, 1957). The segmentation of the germ band in all Opiliones investigated to date, including *P. opilio*, is virtually identical: it consists of the initial formation of the prosomal segments nearly simultaneously and the sequential addition of nine posterior (opisthosomal) segments and telson (Moritz, 1957). This mode of segmentation is conserved in arachnids, including spiders (Akiyama-Oda and Oda, 2020; Anderson, 1973; Damen, 2007; Oda et al., 2007; Paese et al., 2018; Schönauer et al., 2016; Stollewerk et al., 2003a). The initial formation of anterior segments and sequential addition of posterior segments from a growth zone is a feature also observed in Myriapoda (Brena and Akam, 2012) and most insects (Rosenberg et al., 2009; Stahi and Chipman, 2016). At a germ band stage of the spider *P. tepidariorum*, *engrailed-1* stripes of expression in the prosoma form in a rapid sequence: the stripe in L1 appears first, followed by the stripe in the pedipalp segment and in L4, followed by the remaining stripes in the chelicera and other leg bearing segments (Schwager et al., 2009). In *P. opilio*, even though we detected the simultaneous occurrence of six stripes in stage 6 embryos and no expression in stage 5, we cannot rule out the possibility that the temporal resolution of our fixation regime was not adequate to capture the dynamic expression of *Popi-en* in the early germ band. In

contrast to spiders (Anderson, 1973; Mittmann and Wolff, 2012; Wolff and Hilbran, 2011), there are no vestiges of opisthosomal appendages throughout the development of *P. opilio*, consistent with the absence of appendage-like opisthosomal organs in adults of Opiliones (Holm, 1947; Juberthie, 1964; Moritz, 1957).

The timing of limb bud elongation and opisthosomal segmentation is heterochronic in the harvestmen relative to spiders. Whereas limb buds in spiders begin to protrude when three to five opisthosomal segments have formed (Mittmann and Wolff, 2012; Wolff and Hilbran, 2011), the harvestman stage with comparable limb buds coincides with just one opisthosomal segment. Similarly, by the time the harvestman embryo forms five opisthosomal segments, the tips of the legs are touching each other or crossing over the ventral midline and podomere boundaries are clearly visible. This indicates that appendage growth is greatly accelerated in harvestmen with respect to antero-posterior segmentation, in comparison to spiders. The heterochrony observed in these groups could reflect a developmental mechanism to achieve the morphogenesis of elongate appendages. Investigating this hypothesis requires further investigation of other arachnid orders with elongated appendages, such as Amblypygi (whip spiders) and Uropygi (vinegaroons).

Neurogenesis and myogenesis

The first account of embryonic neurogenesis in Opiliones was the description in *Opilio parietinus* of "numerous small pits (…) each one surrounded by a well defined circle of cells" at a stage where the labrum is already formed (Holm, 1947). Each pit is an invagination site of a group of neural precursor cells (Doeffinger et al., 2010; Stollewerk, 2002; Stollewerk et al., 2001; Stollewerk et al., 2003b) that also occur along the ventral ectoderm. In *P. opilio*, invagination sites are first detected at stage 8, prior to the formation of the labrum and when the stomodeum begins to form. This pattern of invagination is typical of arachnids and

myriapods (Dove and Stollewerk, 2003; Stollewerk et al., 2001), in which the neural precursors invaginate as bottle- shaped cells to form the ventral nerve cord and different regions of the brain (Doeffinger et al., 2010; Schwager et al., 2015b; Stollewerk, 2002; Stollewerk et al., 2001; Stollewerk et al., 2003b). In *Drosophila* and in spiders, neural precursors are patterned by Notch signaling, and genes in the *achaete-scute* gene complex (Doeffinger et al., 2010; Stollewerk, 2002; Stollewerk et al., 2001; Stollewerk et al., 2003b), but these genes have not been investigated in Opiliones. A few pleiotropic genes also involved in neural development, such as *orthodenticle*, *Paired box-6*, *empty spiracles*, *dachshund*, and *Distal-less,* have been studied in *P. opilio*, and their expression patterns in the ventral ectoderm and head are comparable to the expression data in other arachnids (Garwood et al., 2014; Sharma et al., 2012a).

Information about embryonic myogenesis of arachnids is very sparse (reviewed by [Anderson, 1973]), and mostly undescribed in Opiliones. In general, the somatic musculature differentiates from mesodermal cells on the appendicular lobes of the paired somites along the body (Anderson, 1973). Winkler (1957) observed paired coelomic sacs on the body segments, but did not mention further the development of the mesoderm. Moritz (1957) described the development of the coxal glands from the somatic mesoderm of limb bud coeloms and depicted proximal appendicular muscle fibers. Using fluorescent antibodies, we confirmed Winkler's observations of the prosomal appendicular lobes of the coelomic sacs and visualized putative muscle precursors in the prosomal appendages and opisthosomal segments, and the formation of the pharyngeal musculature. Considering the scarcity of data on early myogenesis in chelicerates, it would be interesting to investigate the existence of single muscle precursor cells as seen in insects and crustaceans, and test in chelicerates the utility of antibodies developed specifically to target arthropod myoblasts and muscles (Kreissl et al., 2008).

Phalangium opilio as an emerging model for comparative studies of chelicerate development

Despite the last two decades of exciting advances in our knowledge of chelicerate development, there remains challenges to advancing the field of evolutionary developmental biology in Chelicerata for the following reasons. Firstly, classical works on the evolution of development have assumed that the development of horseshoe crabs (Xiphosura), the long-held sister group of Arachnida, reflects ancestral or plesiomorphic traits (Anderson, 1973; Mittmann and Scholtz, 2001). Such evolutionary reconstructions require reinterpretation due to the modern understanding of the phylogenetic position of Xiphosura as nested within Arachnida, a relationship robustly recovered in recent phylogenomic studies (Ballesteros and Sharma, 2019; Sharma et al., 2014c) (but see [Lozano-Fernandez et al., 2019]). More generally, marine arthropod taxa tend to feature larval stages with incomplete segmentation (e.g., sea spiders; horseshoe crabs; marine crustacean groups), whereas several groups of terrestrial arthropods have convergently evolved epimorphic or hemianamorphic development. Secondly, most findings about the molecular mechanisms of development in Arachnida have only been studied in spiders. For instance, it remains unexplored if the cumulus migration mechanism of axis specification, dependent on *decapentaplegic* and *Ets4* signaling (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006; Pechmann et al., 2017) is valid for non-spider arachnids. Thirdly, spiders and other orders of the clade Arachnopulmonata (sensu [Ontano et al., 2021]) have undergone a clade-specific partial- or whole-genome duplication event (Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2014a; Sharma et al., 2015b). Many of these duplicated paralogs appear to have assumed new functions in development or subdivided ancestral functions (Leite et al., 2018). Characterizing the function of single-copy orthologs in arachnids with unduplicated genomes is therefore critical to polarizing gene expression patterns and inferring ancestral states in a phylogenetic context (Nolan et al., 2020). Given our current

understanding of chelicerate phylogeny, unduplicated genomes are the plesiomorphic condition of arachnids (Gainett et al., 2021; Leite et al., 2018; Ontano et al., 2021; Schwager et al., 2017). Therefore, a broader understanding of the development and genomic evolution in chelicerates would greatly benefit from more studies in non-spider (and specifically, non-arachnopulmonate) models.

Encouragingly, non-spider models available include the horseshoe crab *Limulus polyphemus* (Xiphosura), the Arizona bark scorpion *Centruroides sculpturatus* (Scorpiones), the mites *Archegozetes longisetosus* and *Tetranychus urticae* (Acariformes), the tick *Rhipicephalus microplus* (Parasitiformes) (Santos et al., 2013), and more recently, the whip spider *Phrynus marginemaculatus* (Gainett and Sharma, 2020). Among these species, published genomes are available for horseshoe crabs (Battelle et al., 2016; Kenny et al., 2015; Nossa et al., 2014; Shingate et al., 2020a; Shingate et al., 2020b), the mite *T. urticae* (Grbić et al., 2011), and the bark scorpion *C. sculpturatus* (Schwager et al., 2017). While all these species have revealed valuable aspects of the development of their lineages (Barnett and Thomas, 2013; Khila and Grbić, 2007; Sharma et al., 2014a), the autapomorphic body plan of mites, and the current inaccessibility of functional approaches in scorpions (together with their phylogenetic position in the duplicated-genome clade Arachnopulmonata), are disadvantages for their use in comparative studies of chelicerate development. Gene misexpression techniques in horseshoe crabs have also not been achieved. Notably, recent advances in the deployment of CRISPR-mediated mutagenesis have been made in both mite and tick models (Dermauw et al., 2020; Sharma et al., 2020).

Phalangium opilio shares all the desirable qualities of other available non-spider models; two developmental transcriptomes and a draft genome are publicly available (Gainett et al., 2021; Sharma et al., 2013). *P. opilio* is tractable to gene expression essays by whole mount colorimetric in situ hybridization (Sharma et al., 2012b), HCR fluorescent in situ

hybridization (Gainett et al., 2021) and protein expression assays by fluorescent immunochemistry (Sharma, 2018). Robust protocols for embryonic RNAi are available to test gene function (Gainett et al., 2021; Sharma et al., 2013; Sharma et al., 2015a). While the precise phylogenetic position of Opiliones in Arachnida remains contentious (Ballesteros and Sharma, 2019; Lozano-Fernandez et al., 2019; Sharma et al., 2014c), their exclusion from Arachnopulmonata is unambiguous, as well as the unduplicated nature of their genomes (Gainett et al., 2021; Leite et al., 2018). Harvestmen also lack the parasitic lifestyle, miniaturization, and genomic rearrangements exhibited by some mite and tick species, as well as the developmental particularities of scorpions (e.g., ovoviviparity) and horseshoe crabs (e.g., seasonal breeding; anamorphic development in postembryonic stages; threefold whole genome duplication), which incur additional challenges for developmental study. For these reasons, *Phalangium opilio* is poised to serve as a linchpin for comparative studies of chelicerate embryology and genomic evolution across Arthropoda.

Conclusion

The embryonic staging system presented here for the emerging model organism *P. opilio* brings a fresh perspective to the exceptional classical works on the embryology of Opiliones, through the use of molecular biology and modern imaging tools. The staging provides clear landmarks to identify ontogenetic stages and has potential to facilitate collaborative research on this species. The continued research and development of tools to study the common harvestman are expected to clarify aspects of its early embryogenesis, as well as resolve which developmental mechanisms are ancestral to chelicerates and indeed all Arthropoda.

Methods

Embryo cultivation and fixation

Approximately 30 adult individuals of the species *Phalangium opilio* were collected during the months of May through July of 2017 in Madison, Wisconsin (USA). Animals were housed in rectangular plastic containers (25 cm × 12 cm × 12 cm) (Fig. 1C) in groups of 3–4 with at least one male and female, and kept in a room at 24ºC. Each container was provided with coconut fiber substrate, wet cotton to increase humidity, pieces of egg carton for shelter, and egg laying dishes consisting of 35 mm petri dishes with moist coconut fiber (Fig. 1C– E). Animals were fed fish flakes ad libitum, as well as freshly sacrificed *Acheta domestica*. Clutches of 15 to 200 embryos (Fig. 1D–E) laid on petri dishes were transferred to a 26ºC incubator together with large glass beakers containing water to maintain humidity. For incubation, eggs were kept in the moist coconut fiber in petri dishes, or transferred to dampened Whatman paper. For all procedures performed in the embryos, eggs were removed from the coconut fiber using blunt forceps previously touched onto a droplet of water.

For timing the embryonic development, one clutch was selected and fixed at regularly spaced time intervals. From egg laying to 5 days (120h) after egg laying, subsets of embryos were fixed every 12h (11 sampled intervals). From day 5 to day 13, embryos were fixed every 3h (42 sampled intervals), and thereafter every 24h until hatching on day 24 (seven sampled intervals). We followed hatched individuals to adulthood, recording the dates of the molts, and preserving one to two individuals at each molt in 70% ethanol. For approximate times of stages 15–19, eighteen embryos at stage 13 and 14 from a third clutch were dechorionated for 10–20 min in 50% bleach solution and photographed under 1 × phosphate buffered saline (PBS) every day until hatching.

Before fixation, embryos were rinsed in deionized water and dechorionated in 100% bleach solution for 5 min. Embryos were washed in 1 × PBS solution and fixed in the phase between 4% formaldehyde diluted in 1 × PBS supplemented with an equal volume of heptane.

Fixations were carried on a shaker platform overnight (for in situ hybridization) or for a maximum of two hours (for immunohistochemistry). Fixation was stopped by 1 × PBS + 0.02% Tween-20 (Sigma-Aldrich; PBST) washes. Embryos were gradually dehydrated into 96% ethanol and stored at -20ºC.

In situ hybridization

In situ hybridization and probe synthesis was performed according to a modified version of the protocol of Akiyama-Oda and Oda (2003) for *Popi-engrailed* (*en*) and *Popi-vasa*, following (Sharma et al., 2012b). Primers for *Popi-vasa* (forward: 5′-tgcctcctaaaagcgaaaga-3′; reverse: 5′-cat- catccccaaaagaggaa-3′) and *Popi-en* (forward: 5′-cgtc-cgatttttacgttctca-3′; reverse: 5′-cgttaactcctccgttaggc-3′) were designed with T7 linkers for probe synthesis from PCR templates (5′-ggccgcgg-3′ for the forward primer, and 5′-cccggggc-3′ for the reverse). Embryos were imaged using a Nikon SMZ25 fluorescent stereomicroscope equipped with a DS-Fi2 digital color camera, driven by Nikon Elements software.

Immunochemistry

Fluorescent immunohistochemistry was performed according to a modified version of the protocol of Akiyama-Oda & Oda (2003). Vitelline membranes were manually removed with fine forceps. Embryos were rehydrated into 1× PBS + 0.05 Triton X-100 (Sigma-Aldrich; PBS-Tri- ton) and blocked using a 5% normal goat serum + 0.1% bovine serum in PBS-Triton. Primary antibodies consisted of Rat-Tropomyosin (Tm1) (ab50567; abcam) at 1:1500 dilution and acetylated α-tubulin (T6793; Sigma-Aldrich) at 1:500 dilution. Secondary antibodies consisted of goat α rat (Alexa Fluor 488 conjugated; Thermofisher) and goat α mouse Alexa Fluor 488 conjugated; Thermofisher) at 1:200 dilution. After staining, embryos were flat-mounted in 70% glycerol for imaging. Imaging was conducted on a Leica 710 confocal laser

microscope. Maximum intensity Z-projections were generated using Fiji-ImageJ v2.0.0-rc-69/1.52n and channels were merged in Photoshop CS6 or higher by overlapping RGB color channels, or by using the screen layers function. Multi- view stacks were stitched using Zen (Zeiss).

Time-lapse imaging

Two freshly laid egg clutches were kept for three days at 26 ºC incubator prior to preparatory procedures for time- lapse imaging. Eggs were dechorionated in 50% commercial bleach for 15 min, rinsed several times in PBS and glued to plastic petri dishes with double-sided tape. Embryos were then submerged in halocarbon 700 oil (Sigma-Aldrich) for imaging. Imaging was conducted at 20ºC–22 ºC. Images were taken every 15 min on a Nikon SMZ25. Videos of individual embryos have been cropped for clarity. For time-lapse 1 (Additional file 3: File S3), five out of twelve embryos developed normally during the experiment. For time-lapse 2 (Additional file 4: File S4), seven out of ten embryos developed normally during the experiment.

Image editing

Some images have been cropped and/or rotated from original microscope view and were adjusted for brightness, contrast, or levels in Photoshop CS6 or higher, avoiding over-exposing pixels at any point of the micrograph. Hoechst images taken on the Nikon SMZ25 were color-adjusted in Photoshop (hue: -25) for enhanced contrast against black background. Figure plates were assembled using Illustrator CS6 or higher.

Supplementary Information

The online version contains supplementary material available at https://doi.

org/10.1186/s12983-022-00454-z.

Additional file 1: File S1. Oviposition of egg clutch. This video shows a female with protruded ovipositor laying one egg in a clutch deposited in wet cotton.

Additional file 2: File S2. Overview of *P. opilio* development, presenting main features of each stage and approximate timing. Greyscale images were stained with the nuclear marker Hoechst.

Additional file 3: File S3. Time-lapse imaging of embryos of the same clutch 3 days after egg laying. Imaging under oil. Videos have been cropped from the same microscope view. Left: Ventral view; head is to the right. Center: Lateral view; head is to the right. Right: Ventral view; head is to the left. Pictures have been taken every 15 min.

Additional file 4: File S4. Time-lapse imaging of embryos of the same clutch (but different from S1–3) 3 days after egg laying. Imaging under oil. Videos have been cropped from the same microscope view. Left: Dorsal view; head is to the right. Yellow arrowhead at 100.5 h marks the denser spot of cells. Center: Dorsal view, head is facing down. Right: lateral-ventral view; head is facing down. Pictures have been taken every 15 min.

Additional file 5: File S5. Hatching of *Phalangium opilio* into the postembryo. Each frame corresponds to a 10s interval.

Additional file 6: File S6. First molt of *Phalangium opilio*, from postembryo to the first instar. Each frame corresponds to a 10s interval.

Acknowledgements

Microscopy was performed at the Newcomb Imaging Center (Department of Botany, University of Wisconsin-Madison). We are indebted to Sarah Swanson for facilitating training and use of the imaging facility. Jesús Ballesteros, Andrew Z. Ontano, Carlos E. Santibáñez López, and Rachel M. Smaby assisted with collection of *P. opilio*. Comments from the

Associate Editor Angelika Stollewerk, Seth Donoughe, and one anonymous reviewer improved a previous draft of the work.

Authors' contributions

ARC and CS conducted serial fixations of synchronous clutches. BCK timed and live-imaged late embryonic stages. CMB and EVWS provided photographs and videos of adult harvestmen and laboratory setup. ARC and GG conducted in situ hybridization and imaging. GG and PPS conducted immunohistochemistry and confocal microscopy. GG and PPS analyzed the data, assembled figures and drafted the manuscript. PPS conceived the study. All authors assisted with colony maintenance. All authors read and approved the final manuscript.

Funding

This work was supported by the National Science Foundation grant no. IOS- 1552610 and IOS-2016141.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its Additional files].

Declarations

Ethics approval and consent to participate

Research using *P. opilio* is not subject to approval from an ethics committee.

Consent for publication

Not applicable.

Competing interests

The authors declare having no conflict of interest.

Received: 2 November 2021 Accepted: 9 February 2022

Published online: 04 March 2022

Footnotes

[1]In the published version of this chapter, we used the species name *Odiellus gallicus*, which is not a valid species name. *Acantholophus gallicus* Simon 1879 is the protonym, which was synonymized by Šilhavý (1949a: 11) with *Phalangium troguloides* Lucas, 1847. Roewer (1923) had transferred *Phalangium troguloides* Lucas, 1847 to *Odiellus troguloides*. Therefore, from 1949 onwards, *Odiellus troguloides* Lucas, 1847 is the most recent valid name of *Acantholophus gallicus* Simon 1879.

References

Abzhanov, A., Popadić, A. and Kaufman, T. C. (1999). Chelicerate Hox genes and the homology of arthropod segments. *Evol. Dev.* 1, 77–89.

Akiyama-Oda, Y. and Oda, H. (2003). Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* 130, 1735–1747.

Akiyama-Oda, Y. and Oda, H. (2006). Axis specification in the spider embryo: *dpp* is required for radial-to-axial symmetry transformation and *sog* for ventral patterning. *Development* 133, 2347–2357.

Akiyama-Oda, Y. and Oda, H. (2010). Cell migration that orients the dorsoventral axis is coordinated with anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. *Development* 137, 1263–1273.

Akiyama-Oda, Y. and Oda, H. (2020). Hedgehog signaling controls segmentation dynamics and diversity via *msx1* in a spider embryo. *Sci. Adv.* 6, eaba7261.

Anderson, D. T. (1973). Chelicerates. In *Embryology and Phylogeny of Annelids and Arthropods*, pp. 365–451. Pergamon.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440–22.

Barnett, A. A. and Thomas, R. H. (2012). The delineation of the fourth walking leg segment is temporally linked to posterior segmentation in the mite *Archegozetes longisetosus* (Acari: Oribatida, Trhypochthoniidae). *Evol. Dev.* 14, 383–392.

Barnett, A. A. and Thomas, R. H. (2013). Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *Evodevo* 4, 23.

Battelle, B.-A., Ryan, J. F., Kempler, K. E., Saraf, S. R., Marten, C. E., Warren, W. C., Minx, P. J., Montague, M. J., Green, P. J., Schmidt, S. A., et al. (2016). Opsin repertoire and expression patterns in horseshoe crabs: Evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biol. Evol.* 8, 1571–1589.

Baudouin-Gonzalez, L., Schoenauer, A., Harper, A., Blakeley, G., Seiter, M., Arif, S., Sumner-Rooney, L., Russell, S., Sharma, P. P. and McGregor, A. P. (2021). The evolution of Sox gene repertoires and regulation of segmentation in arachnids. *Mol. Biol. Evol*. 38, 3153–3169.

Brena, C. and Akam, M. (2012). The embryonic development of the centipede *Strigamia maritima*. *Dev. Biol.* 363, 290–307.

Buzatto, B. A. and Machado, G. (2014). Male dimorphism and alternative reproductive tactics in harvestmen (Arachnida: Opiliones). *Behav. Proc.* 109, 2–13.

Chipman, A. D. (2010). Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. *BioEssays* 32, 60–70.

Damen, W. G. M. (2007). Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev. Dyn.* 236, 1379–1391.

Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. and Tautz, D. (1998). A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl. Acad. Sci. USA* 95, 10665–10670.

Dermauw, W., Jonckheere, W., Riga, M., Livadaras, I., Vontas, J. and Leeuwen, T. Van (2020). Targeted mutagenesis using CRISPR-Cas9 in the chelicerate herbivore *Tetranychus urticae*. *Insect Biochem. Mol. Biol.* 120, 103347.

Doeffinger, C., Hartenstein, V. and Stollewerk, A. (2010). Compartmentalisation of the

precheliceral neuroectoderm in the spider *Cupiennius salei*: Development of the arcuate body, the optic ganglia and the mushroom body. *J. Comp. Neurol.* 518, 2612–2632.

Dove, H. and Stollewerk, A. (2003). Comparative analysis of neurogenesis in the myriapod *Glomeris marginata* (Diplopoda) suggests more similarities to chelicerates than to insects. *Development* 130, 2161–2171.

Edgar, A., Bates, C., Larkin, K. and Black, S. (2015). Gastrulation occurs in multiple phases at two distinct sites in *Latrodectus* and *Cheiracanthium* spiders. *Evodevo* 6, 33.

Faussek, V (1891). Zur Anatomie und Embryologie der Phalangiden. *Trav. Soc. Imp. Natur. St. Pétersb.* 22, 2.

Gainett, G. and Sharma, P. P. (2020). Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *Evodevo* 11, 18.

Gainett, G., Michalik, P., Müller, C. H. G., Giribet, G., Talarico, G. and Willemart, R. H. (2017a). Putative thermo-/hygroreceptive tarsal sensilla on the sensory legs of an armored harvestman (Arachnida, Opiliones). *Zool. Anz.* 270, 81–97.

Gainett, G., Michalik, P., Müller, C. H. G., Giribet, G., Talarico, G. and Willemart, R. H. (2017b). Ultrastructure of chemoreceptive tarsal sensilla in an armored harvestman and evidence of olfaction across Laniatores (Arachnida, Opiliones). *Arthropod Struct. Dev.* 46, 178–195.

Gainett, G., González, V. L., Ballesteros, J. A., Setton, E. V. W., Baker, C. M., Gargiulo, L. B., Santibáñez-López, C. E., Coddington, J. A. and Sharma, P. P. (2021). The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proc. R. Soc. B* 288, 20211168.

Garwood, R. J., Sharma, P. P., Dunlop, J. A. and Giribet, G. (2014). A Paleozoic stem group to mite harvestmen revealed through integration of phylogenetics and development. *Curr. Biol.* 24, 1017–1023.

Giribet, G. and Sharma, P. P. (2015). Evolutionary biology of harvestmen (Arachnida, Opiliones). *Annu. Rev. Entomol.* 60, 157–175.

Gnaspini, P. (2007). Development. In *Harvestmen: The Biology of Opiliones*. ed. Pinto-da-Rocha, R., Machado, G., and Giribet, G., pp. 455–472. Cambridge: Harvard University Press.

Gnaspini, P. and Hara, M. (2007). Defense mechanisms. In *Harvestmen: The Biology of Opiliones*. ed. Pinto-da-Rocha, R., Machado, G., and Giribet, G., pp. 374–99. Cambridge: Harvard University Press.

Grbić, M., Leeuwen, T. Van, Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.

Holm, A. (1947). On the development of *Opilio parietinus* Deg. *Zoologiska Bidrag Fran Uppsala* 1–17.

Holm, A (1952). Experimentelle Untersuchungen über die Entwicklung und Entwicklungsphysiologie des Spinnenembryos. *Zool. Bidr. Uppsala* 29, 293–424.

Juberthie, C. (1961). Les phases du développement embryonnaire et leurs relations avec la température et l'humidité chez un Opilion Palpatores. *C. R. Biol.* 252, 2144–2142.

Juberthie, C. (1964). *Recherches sur la biologie des Opilions*. Paris: Centre national de la recherche scientifique.

Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., et al. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116, 190–199.

Khila, A. and Grbić, M. (2007). Gene silencing in the spider mite *Tetranychus urticae*:

dsRNA and siRNA parental silencing of the *Distal-less* gene. *Dev. Genes. Evol.* 217, 241–251.

Kreissl, S., Uber, A. and Harzsch, S. (2008). Muscle precursor cells in the developing limbs of two isopods (Crustacea, Peracarida): an immunohistochemical study using a novel monoclonal antibody against myosin heavy chain. *Dev. Genes. Evol.* 218, 253–265.

Kury, A. B., Mendes, A. C., Cardoso, L., Kury, M. S., Granado, A. A., Yoder, M. J. and Kury, I. S. (2021). WCO-Lite version 1.1: an online nomenclatural catalogue of harvestmen of the world (Arachnida, Opiliones) curated in TaxonWorks. *Zootaxa* 4908, 447–450.

Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N.-M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018). Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* 35, 2240–2253.

Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D. and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* 10, 2295–8.

Machado, G. and Macías-Ordónéz, R. (2007). Reproduction. In *Harvestmen: The Biology of Opiliones*. ed. Pinto-da-Rocha, R., Machado, G., and Giribet, G., pp. 414–454. Cambridge: Harvard University Press.

March, L. E., Smaby, R. M., Setton, E. V. W. and Sharma, P. P. (2018). The evolution of selector gene function: Expression dynamics and regulatory interactions of *tiptop*/*teashirt* across Arthropoda. *Evol. Dev.* 20, 219–232.

Mittmann, B. and Scholtz, G. (2001). *Distal-less* expression in embryos of *Limulus polyphemus* (Chelicerata, Xiphosura) and *Lepisma saccharina* (Insecta, Zygentoma) suggests a role in the development of mechanoreceptors, chemoreceptors, and the CNS. *Dev. Genes. Evol.* 211, 232–243.

Mittmann, B. and Wolff, C. (2012). Embryonic development and staging of the cobweb spider *Parasteatoda tepidariorum* C. L. Koch, 1841 (syn.: *Achaearanea tepidariorum*; Araneomorphae; Theridiidae). *Dev. Genes. Evol.* 222, 189–216.

Moritz, M. (1957). Zur Embryonalentwicklung der Phalangiiden (Opiliones, Palpatores) unter besonderer Berücksichtigung der äußeren Morphologie, der Bildung des Mitteldarmes und der Genitalanlage. *Zool. Jahrb. Abt. Anat. Ontog. Tiere* 76, 331–370.

Moritz, M. (1959). Zur Embryonalentwicklungs der Phalangiiden (Opiliones; Palpatores) II. Die Anlage und Entwicklung der Coxaldrüse bei *Phalangium opilio* L.). *Zool. Jahrb. Abt. Anat. Ontog. Tiere* 77, 229–240.

Muñoz-Cuevas, A. (1971). Etude du développement embryonnaire de *Pachylus quinamavidensis* (Arachnides, Opilions, Gonyleptidae). *Bulletin du Muséum national d'histoire naturelle* 6, 1238–1250.

Nolan, E. D., López, C. E. S. and Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Dev. Genes Evol.* 230, 137–153.

Nossa, C. W., Havlak, P., Yue, J.-X., Lv, J., Vincent, K. Y., Brockmann, H. J. and Putnam, N. H. (2014). Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* 3, 708–21.

Oda, H., Nishimura, O., Hirao, Y., Tarui, H., Agata, K. and Akiyama-Oda, Y. (2007). Progressive activation of Delta-Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider *Achaearanea tepidariorum*. *Development* 134, 2195–2205.

Ontano, A. Z., Gainett, G., Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., et al. (2021). Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol. Biol. Evol.* 38, 2446–2467.

Paese, C. L. B., Schoenauer, A., Leite, D. J., Russell, S. and McGregor, A. P. (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *Elife* 7, 1735.

Pechmann, M. (2020). Embryonic development and secondary axis induction in the Brazilian white knee tarantula *Acanthoscurria geniculata*, C. L. Koch, 1841 (Araneae; Mygalomorphae; Theraphosidae). *Dev. Genes Evol.* 1–20.

Pechmann, M., Schwager, E. E., Turetzek, N. and Prpic, N.-M. (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc R. Soc. B.* 282, 20151162.

Pechmann, M., Benton, M. A., Kenny, N. J., Posnien, N. and Roth, S. (2017). A novel role for *Ets4* in axis specification and cell migration in the spider *Parasteatoda tepidariorum*. *Elife* 6, e27590.

Rosenberg, M. I., Lynch, J. A. and Desplan, C. (2009). Heads and tails: Evolution of antero-posterior patterning in insects. *Biochimica Et Biophysica Acta Bba. Gene Regul. Mech.* 1789, 333–342.

Santos, V. T., Ribeiro, L., Fraga, A., Barros, C. M., Campos, E., Moraes, J., Fontenele, M. R., Araújo, H. M., Feitosa, N. M., Logullo, C., et al. (2013). The embryogenesis of the tick *Rhipicephalus* (*Boophilus*) *microplus*: the establishment of a new chelicerate model system. *Genesis* 51, 803–818.

Schimkewitsch W (1898). Ueber die Entwicklung des Darmkanals bei einigen Arachniden. *Trav. Soc. Imp. Natur. St. Pétersb.* 2:25

Schönauer, A., Paese, C. L. B., Hilbrant, M., Leite, D. J., Schwager, E. E., Feitosa, N. M., Eibner, C., Damen, W. G. M. and McGregor, A. P. (2016). The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* 143, 2455–2463.

Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P. and Damen, W. G. M. (2009). *hunchback* functions as a segmentation gene in the spider *Achaearanea tepidariorum*. *Curr. Biol.* 19, 1333–1340.

Schwager, E. E., Meng, Y. and Extavour, C. (2015a). *vasa* and *piwi* are required for mitotic integrity in early embryogenesis in the spider *Parasteatoda tepidariorum*. *Dev. Biol.* 402, 276–290.

Schwager, E. E., Schönauer, A., Leite, D. J., Sharma, P. P. and McGregor, A. P. (2015b). Chelicerata. In *Evolutionary Developmental Biology of Invertebrates 3* (ed. Wanninger, A.), pp. 99–139. Springer, Vienna.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.

Seaver EC. (2003). Segmentation: mono- or polyphyletic? *Int. J. Dev. Biol.* 47, 583–595.

Setton, E. V. W. and Sharma, P. P. (2018). Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proc. Natl. Acad. Sci. USA* 128, 201720193–10.

Setton, E. V. W., March, L. E., Nolan, E. D., Jones, T. E., Cho, H., Wheeler, W. C., Extavour, C. G. and Sharma, P. P. (2017). Expression and function of *spineless* orthologs correlate with distal deutocerebral appendage morphology across

Arthropoda. *Dev. Biol.* 430, 224–236.
Sharma, P. P. (2017). Chelicerates and the conquest of land: a view of arachnid origins through an evo-devo spyglass. *Integr. Comp. Biol.* 57, 510–522.
Sharma, P. P. (2018). Chelicerates. *Curr. Biol.* 28, R774–R778.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012a). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* 14, 522–533.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012b). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.
Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol. Dev.* 15, 228–242.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Wheeler, W. C. (2014a). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661–20140661.
Sharma, P. P., Gupta, T., Schwager, E. E., Wheeler, W. C. and Extavour, C. G. (2014b). Subdivision of arthropod *cap-n-collar* expression domains is restricted to Mandibulata. *Evodevo* 5, 3.
Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C. and Giribet, G. (2014c). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31, 2963–2984.
Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C. and Extavour, C. G. (2015a). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. B* 282, 20150698.
Sharma, P. P., Santiago, M. A., González-Santillán, E., Monod, L. and Wheeler, W. C. (2015b). Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol. Dev.* 17, 347–355.
Sharma, A., Pham, M. N., Reyes, J. B., Chana, R., Yim, W. C., Heu, C. C., Kim, D., Chaverra-Rodriguez, D., Rasgon, J. L., Harrell, R. A., et al. (2020). Cas9-mediated gene-editing in the black-legged tick, *Ixodes Scapularis*, by embryo injection and ReMOT Control. *Ssrn Electron J.*
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E. and Venkatesh, B. (2020a). Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.* 11, 2322.
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H. and Venkatesh, B. (2020b). Chromosome-level genome assembly of the coastal horseshoe crab (*Tachypleus gigas*). *Mol. Ecol. Resour.* 20, 1748–1760.
Stahi, R. and Chipman, A. D. (2016). Blastoderm segmentation in *Oncopeltus fasciatus* and the evolution of insect segmentation mechanisms. *Proc. R. Soc. B* 283, 20161745.
Stollewerk, A. (2002). Recruitment of cell groups through Delta/Notch signalling during spider neurogenesis. *Development* 129, 5339–5348.
Stollewerk, A., Weller, M. and Tautz, D. (2001). Neurogenesis in the spider *Cupiennius salei*. *Development* 128, 2673–2688.
Stollewerk, A., Schoppmeier, M. and Damen, W. G. M. (2003a). Involvement of *Notch* and *Delta* genes in spider segmentation. *Nature* 423, 863–865.
Stollewerk, A., Tautz, D. and Weller, M. (2003b). Neurogenesis in the spider: new insights from comparative analysis of morphological processes and gene expression patterns.

Arthropod Struct. Dev. 32, 5–16.
Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl. Acad. Sci. USA* 95, 10671–10675.
Willemart, R. H., Farine, J.-P. and Gnaspini, P. (2009). Sensory biology of Phalangida harvestmen (Arachnida, Opiliones): a review, with new morphological data on 18 species. *Acta Zoologica* 90, 209–227.
Winkler, D. (1957). Die Entwicklung Der Äusseren Körpergestalt bei Den Phalangiidae (Opiliones). *Mitt Zool. Mus. Berl.* 33, 355–389.
Wolff, C. and Hilbran, M. (2011). The embryonic development of the central American wandering spider *Cupiennius salei*. *Front. Zool.* 8, 15.
Yoshikura, M. (1975). Comparative embryology and phylogeny of Arachnida. *Kumamoto J. Sci. Ser. B. Sect. Biol.* 12, 71–142.

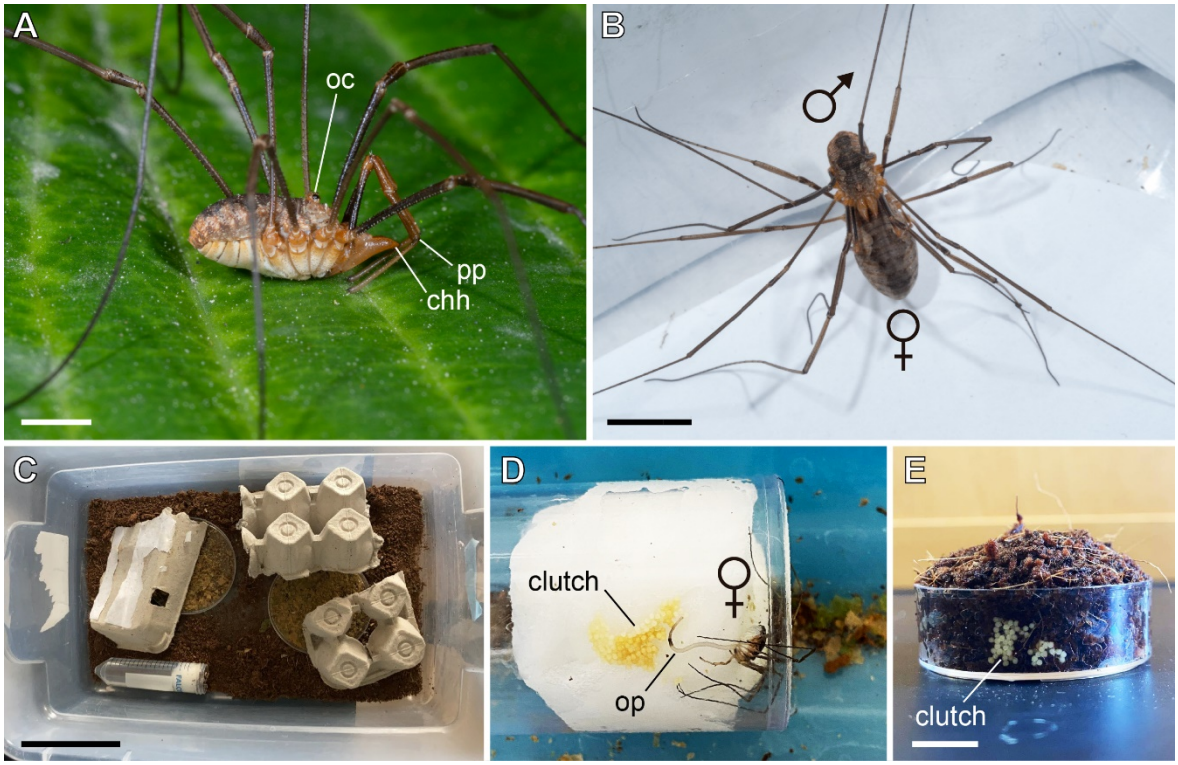


Figure 1: *Phalangium opilio* as a model for developmental biology. **A**: Adult male *P. opilio* in lateral view. Anterior is to the right. **B**: Mating couple of *P. opilio*. **C:** Example of a terraria where adults *P. opilio* are kept at the Sharma lab (UW-Madison). **D**: Adult female laying an egg in a clutch in dampened cotton (water source). **E**: Egg-laying dish, consisting of a 35 mm plastic dish with dampened coconut fiber. Note a clutch on the lateral wall of the dish. chh: cheliceral horn; oc: ocularium; op: ovipositor; pp: pedipalp. Scale bar (approximate values): **A**: 2 mm; **B**: 6 mm; **C**: 10 cm; **E**: 1 cm.

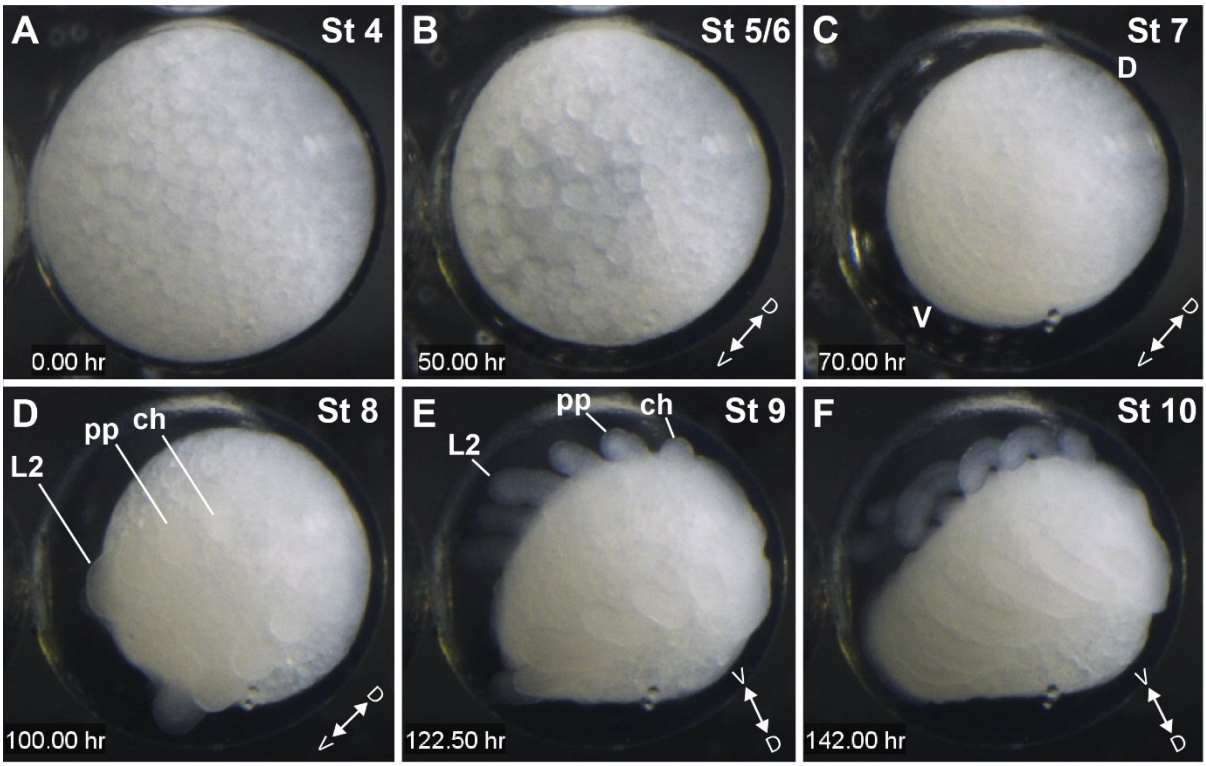


Figure 2: Series of selected frames from time-lapse 1 (Additional file 3: File S3) of an embryo imaged under halocarbon oil at room temperature (20 °C). Photographs were taken every 15 min. **A**: Stage 4, three-day old embryo, at the onset of perivitelline space formation. **B**: Stage 5/6, undergoing cell movements. **C**: Stage 7, with a ventral segmented germ band. **D**: Stage 8, limb bud protrusion. **E**: Stage 9, appendage elongation and further formation of opisthosomal segments**. F**: Stage 10: Ventral inflexion of the opisthosoma. The diameter of the egg is approximately 550 µm. Timing of development is slower than the staging given the cooler temperature at which the embryo was imaged. D and V: dorso-ventral axis; ch: cheliceral segment; L2: leg 2 segment; pp: pedipalpal segment.

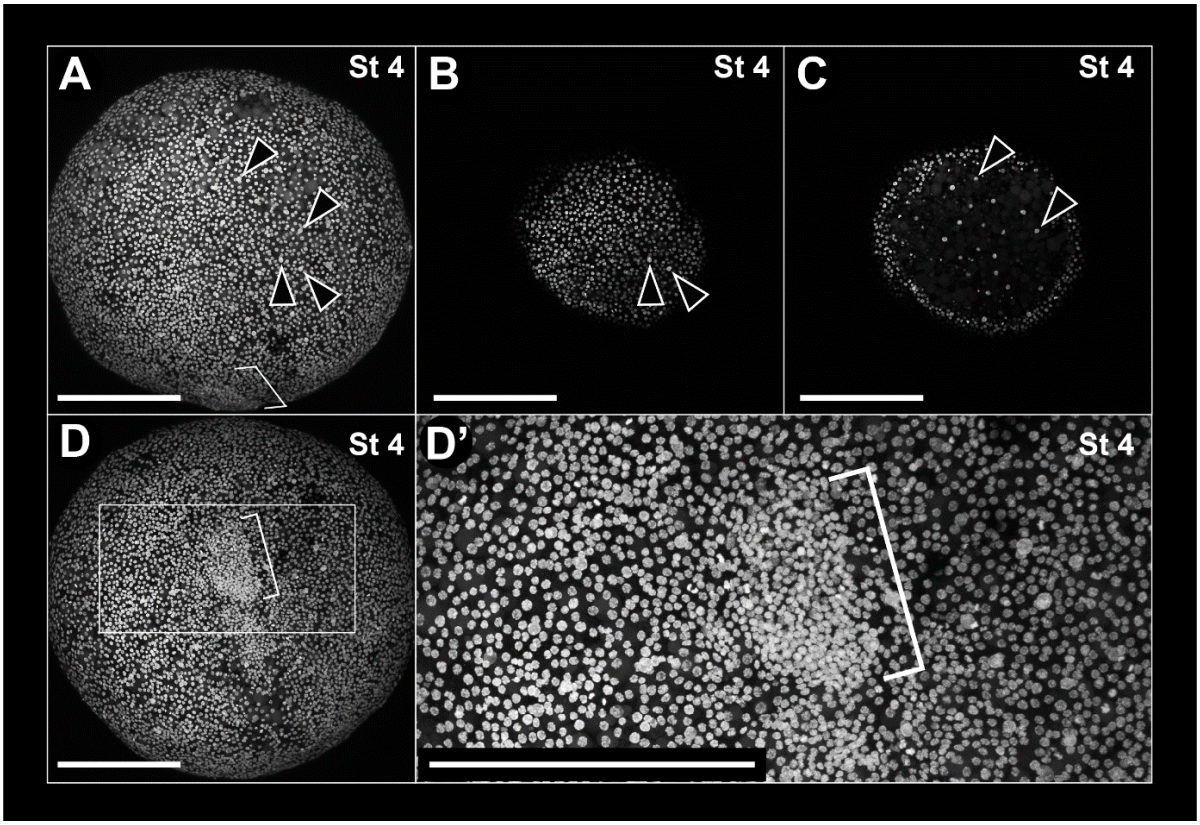


Figure 3: Stage 4 (**A–D′**). Confocal micrographs of an embryo of *P. opilio* stained with Hoechst (nuclei). **A**: Z-stack maximum intensity projection of the whole embryo**. B, C**: Optical sections at 6 and 11 µm depths (outward to inward), respectively. Note that larger nuclei occur more internally to the small nuclei. **D**: Z-stack maximum intensity projection of surface of the embryo shown in "**A**" rotated 90° forward. **D′**: detail of the same image in **D**, showing a denser spot of small nuclei at the surface of the embryo (bracket). Arrowheads: larger sub-peripheral nuclei. Scale bar **A–D′:** 200 µm.

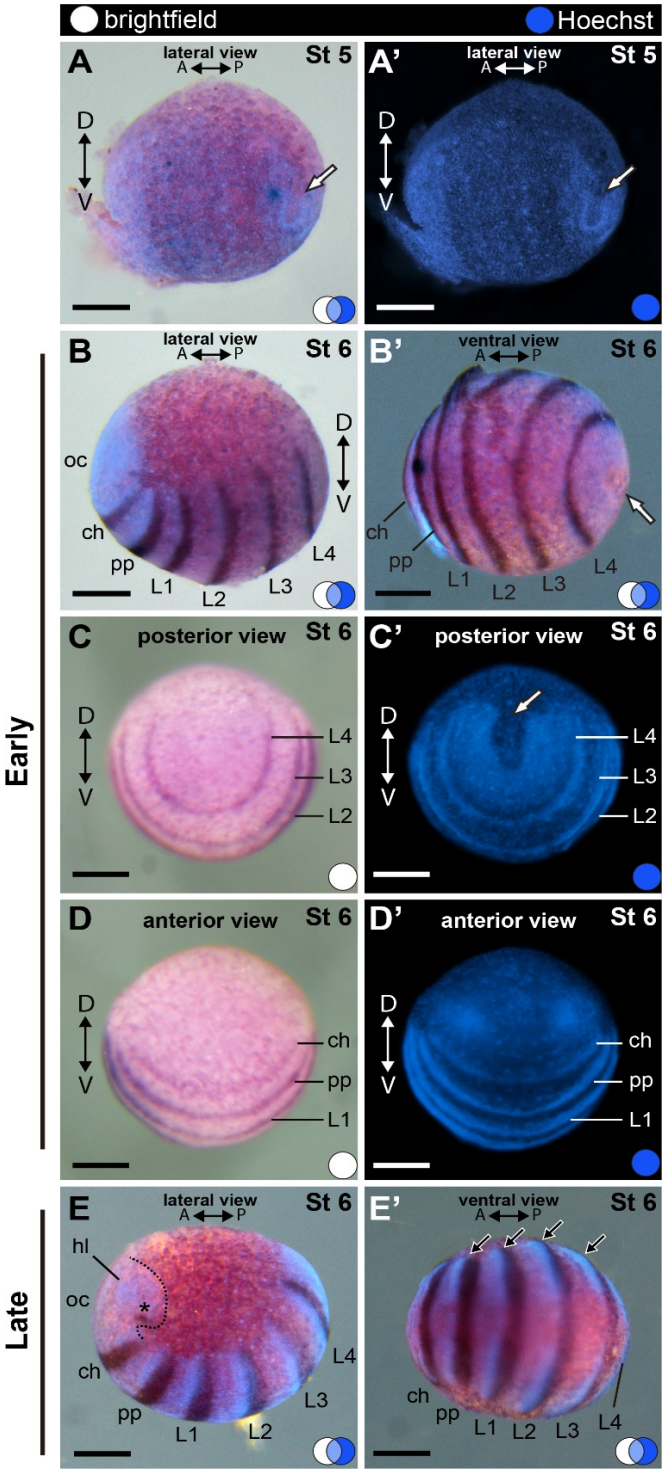


Figure 4: Stage 5 **(A)** and early **(B–D′)** to late **(E–E′)** stage 6. In situ hybridization for *Popi-en*. **A, A′**: Early germ band. No clear expression of *Popi-en* was detected**. B–D′**: Initial segmentation of the prosoma. **D–E′**: C-shaped segmented germ band longer in the A–P axis and with more defined head-lobe. **A** and **P**: anterior–posterior axis; asterisk: dot of engrailed expression; **D** and **V**: dorsal–ventral axis; white arrow: horseshoe shaped posterior end of the germ band; blue circle: Hoechst staining. white circle: bright feld; ch: cheliceral segment; L1–4: leg 1–4 segments; oc: ocular segment; pp: pedipalpal segment. Scale bars: 100 µm.

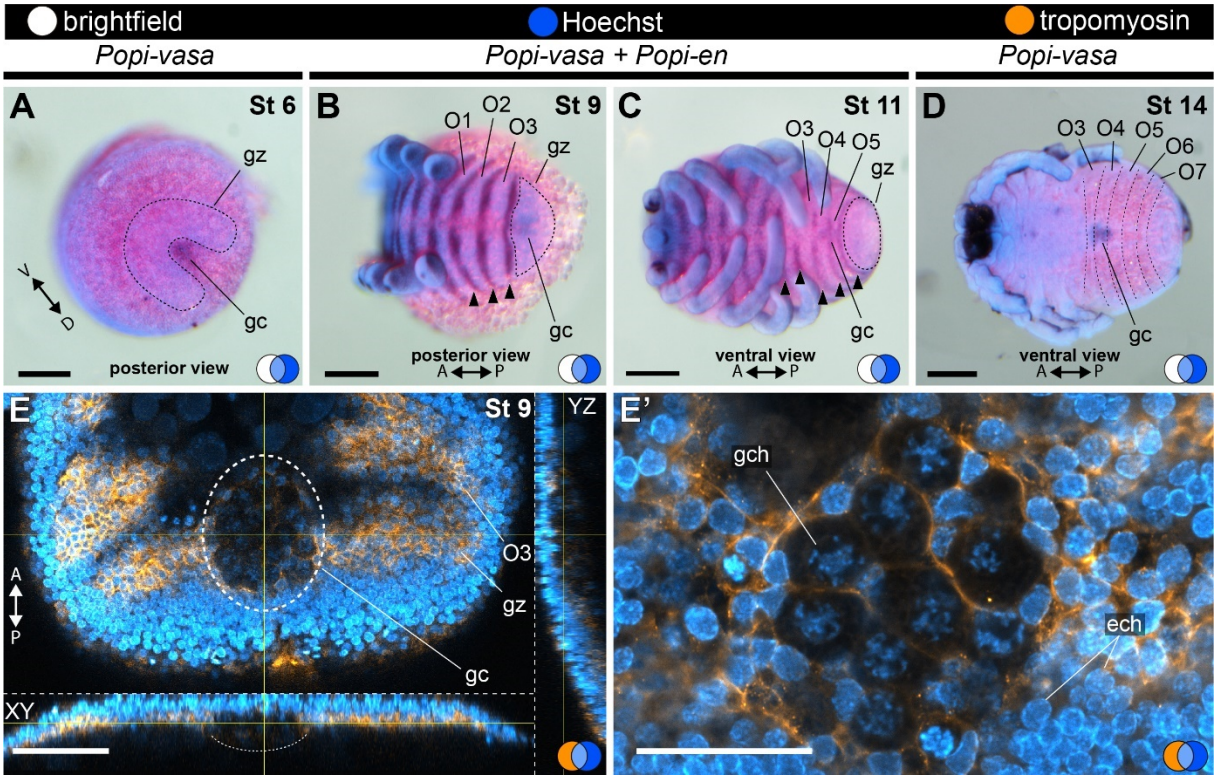


Figure 5: Germ cell development in *P. opilio*. **A, D**: In situ hybridization for *Popi-vasa*. **B–C**: Double in situ hybridization for *Popi-vasa* and *Popi-en*. **E–E′**: Antibody staining for tropomyosin. **A:** Stage 6 embryo. **B**: Stage 9 embryo. **C**: Stage 11 embryo. **D**: Stage 14 embryo**. E**: Single optical slice through the germ cell cluster at stage 9, with orthogonal projections. **E′**: Detail of the germ cell nuclei shown in "**E**". A and P: antero-posterior axis; blue circle: Hoechst; orange circle: tropomyosin; white circle: bright field; ech: ectodermal cell chromatin; gc: germ cell cluster; gch: germ cell chromatin; gz: growth zone; O3–O7: opisthosomal segments 3–7. Scale bar **A–D**: 100 µm**; E**: 50 µm; **E′**: 25 µm.

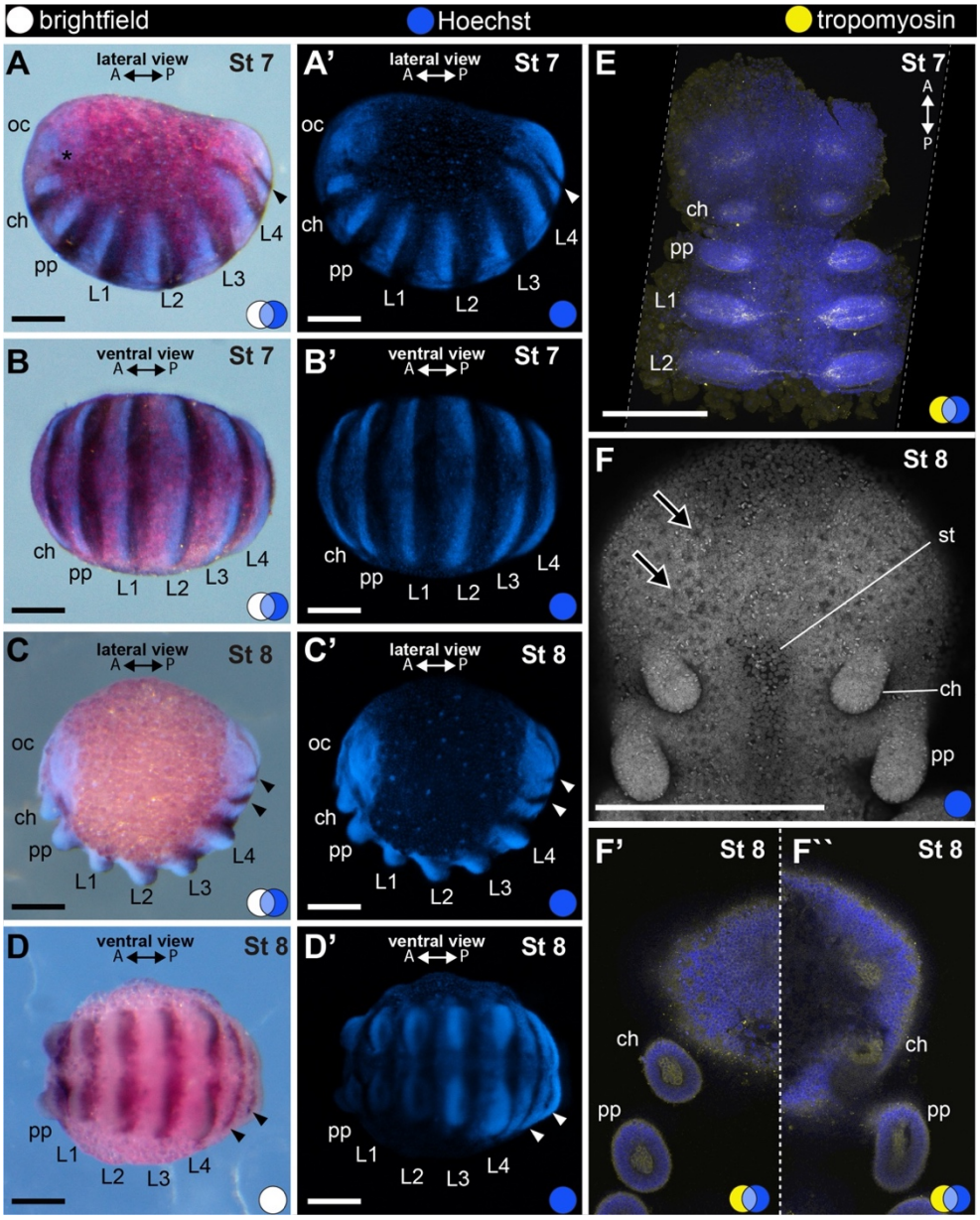


Figure 6: Stage 7 **(A–B′, E)** and stage 8 (**C–D′, F–F′′**). In situ hybridization for the gene *Popi-engrailed* **(A–D′)** and confocal micrographs of fluorescent antibody staining against tropomyosin **(E–F′′)**. Images with the same letters are different views of the same embryo. **A–B′**: First opisthosomal segment. **C–D′**: Second opisthosomal segment. **E**: Anterior half of a fat-mounted germ band (ventral view). Tropomyosin-positive cells occur at the head lobe and primordial limb buds. **F**: Anterior third of a fat-mounted germ band at stage 8 (ventral view). Stomodeum begins to form between cheliceral limb buds**. F′, F′′:** Optical sections at, respectively, 16 and 27 µm depth from the ventral surface of the same sample in **F**, showing tropomyosin-positive cells internal to each limb bud, and a coelomic space. A and P: anterior–posterior axis; arrowhead: stripe of engrailed expression marking the posterior border of opisthosomal segment; asterisk: dot of engrailed expression; black arrow: invagination site of neural precursor cells; blue circle: Hoechst (blue); white circle: bright field; yellow circle: tropomyosin (yellow); ch: cheliceral segment; L1–4: leg 1–4 segments; oc: ocular segment; pp: pedipalpal segment; st: stomodeum. Scale bars A–D′: 100 µm. E–F′′: 250 µm.

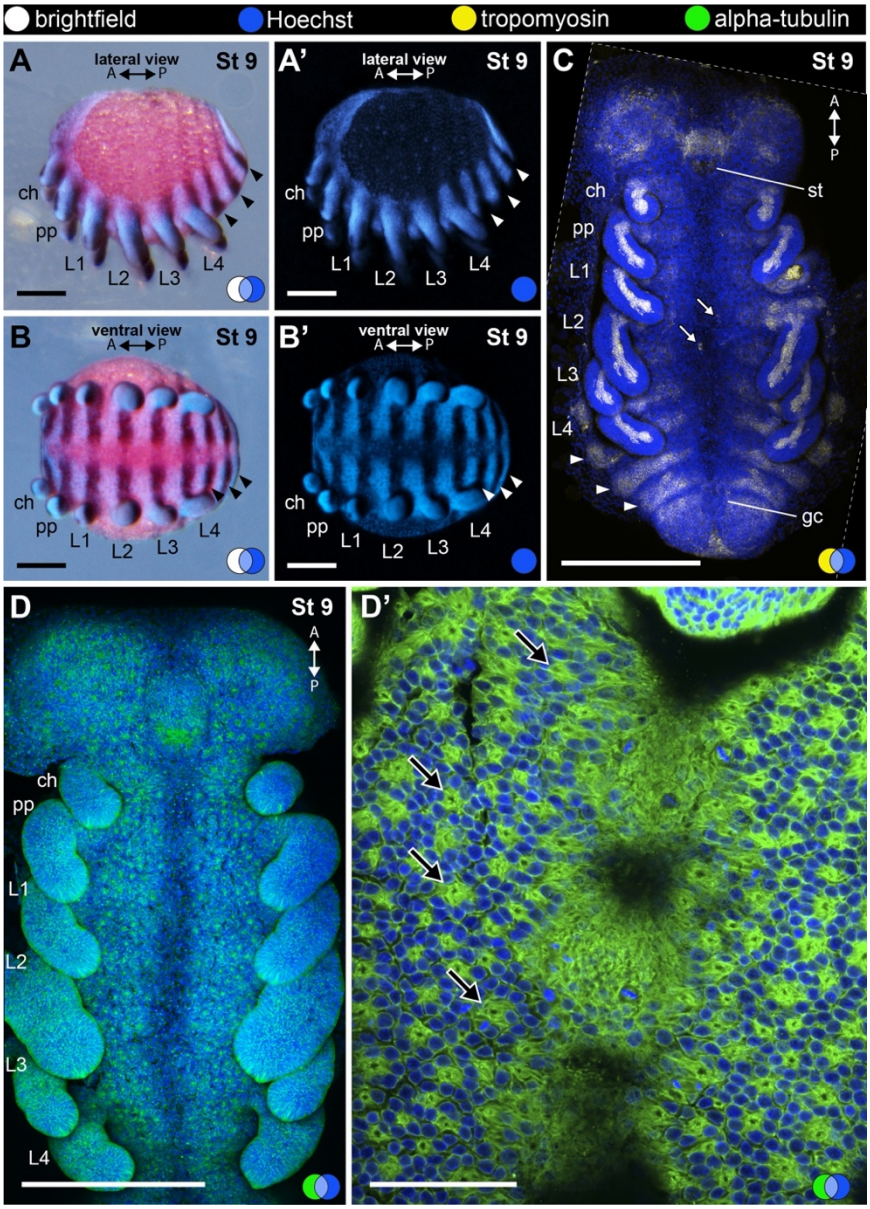


Figure 7: Stage 9 **(A–D′)**. In situ hybridization for the gene *Popi-engrailed* (**A–B′**) and confocal micrographs of fluorescent antibody staining against tropomyosin **(C)** and alpha-tubulin (**D, D′**). Images with the same letters are different views of the same embryo. **A–B′**: Third opisthosomal segment. **C**: Maximum projection of a fat-mounted embryo (ventral view). Tropomyosin-positive cells occur as a bundle inside each prosomal limb buds and in broad stripes in each opisthosomal segment and posterior growth zone. **D**: Maximum projection of a fat-mounted embryo (ventral view). Note groups of invaginating cells along ventral ectoderm and head lobes**. D′**: Detail of a group of invaginating neural precursor cells on the ventral ectoderm. A and P: anterior–posterior axis; arrowhead: stripe of engrailed expression marking the posterior border of opisthosomal segment; black arrow: neural precursor cells (apical process); blue circle: Hoechst (blue); white arrow: isolated tropomyosin-positive cell; white circle: bright field; yellow circle: tropomyosin (yellow); green circle: alpha-tubulin (green); ch: cheliceral segment; gc: germ cell cluster; L1–4: leg 1–4 segments; pp: pedipalpal segment; st: stomodeum. Scale bars **A–B′:** 100 µm. **C, E**: 250 µm. **D** reprinted with permission from (Sharma, 2018).

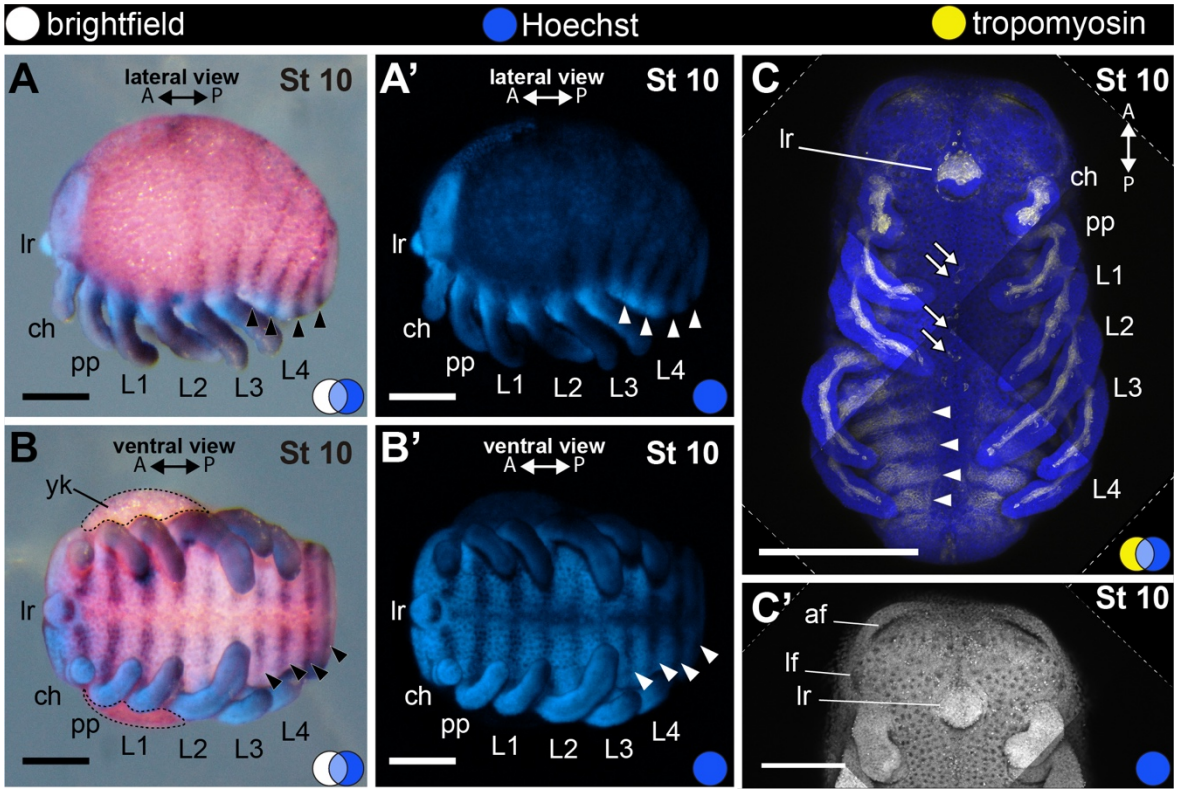


Figure 8: Stage 10 (**A–C′**). In situ hybridization for the gene *Popi-engrailed* (**A–B′**) and confocal micrographs of fluorescent antibody staining against tropomyosin (**C, C′**). **A–B′**: Fourth opisthosomal segment and ventral flexure of the opisthosoma. **C**: Maximum projection of a fat-mounted embryo (ventral view). **C′**: Detail of the head lobes of same preparation shown in **C**. Only Hoechst channel in shown (gray). A and P: anterior–posterior axis; arrowhead: stripe of engrailed expression marking the posterior border of opisthosomal segment; blue circle: Hoechst (blue, gray); white arrow: isolated tropomyosin-positive cell; white circle: bright field; yellow circle: tropomyosin (yellow); af: anterior furrow; ch: cheliceral segment; lr: labrum; lf: lateral furrow; L1–4: leg 1–4 segments; pp: pedipalpal segment; yk: yolk. Scale bars A–B′: 100 µm. C: 250 µm. C′: 125 µm.

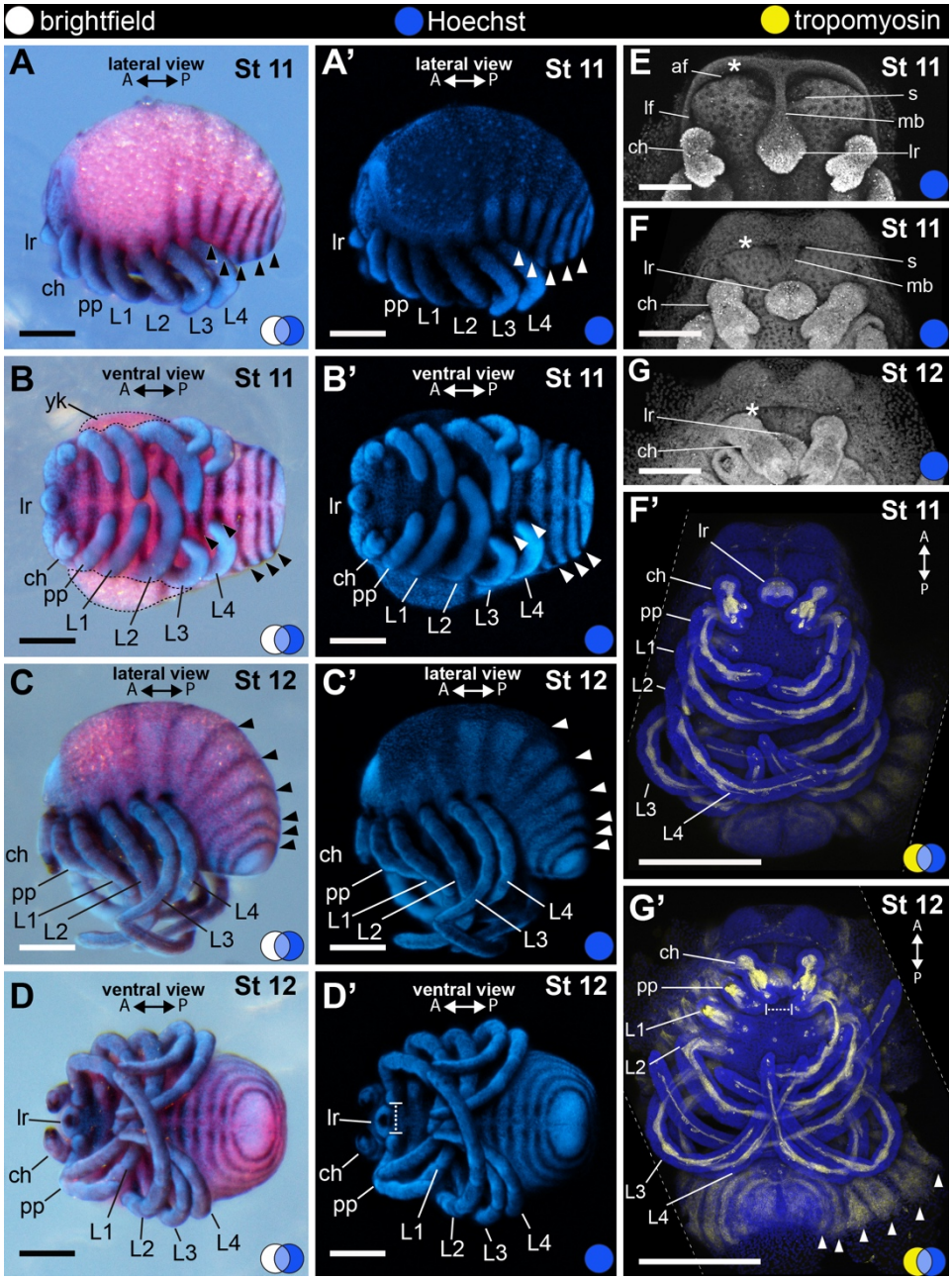


Figure 9: Stage 11 (**A–B′, E–F′**) and stage 12 (**C–D′, G–G′**). In situ hybridization for *Popi-en* (**A–D′**) and confocal micrographs of fluorescent antibody staining against tropomyosin (**F′, G′**). **A–B′:** Fifth opisthosomal segment and initial folding of the anterior furrow**. C–D′**: Sixth opisthosomal segment. **E:** Detail of the head lobes of early stage 11. **F**: Detail of the head lobes of late stage 11 embryo. **F′**: Maximum projection of a fat-mounted late stage 11 embryo (ventral view). **G**: Detail of the head lobes of a stage 12 embryo. **G′**: Maximum projection of a fat-mounted embryo at stage 12 (ventral view). A and P: anterior–posterior axis; arrowhead: stripe of engrailed expression marking the posterior border of opisthosomal segment; blue circle: Hoechst (blue, gray); white circle: bright field; yellow circle: tropomyosin (yellow); asterisk: anterior rim of the folding ectoderm; af: anterior furrow; ch: cheliceral segment; lr: labrum; lf: lateral furrow; L1–4: leg 1–4 segments; mb: medial bridge; pp: pedipalpal segment; s: slit; yk: yolk. Scale bars **A–G**: 100 µm. **F′, G′**: 250 µm.

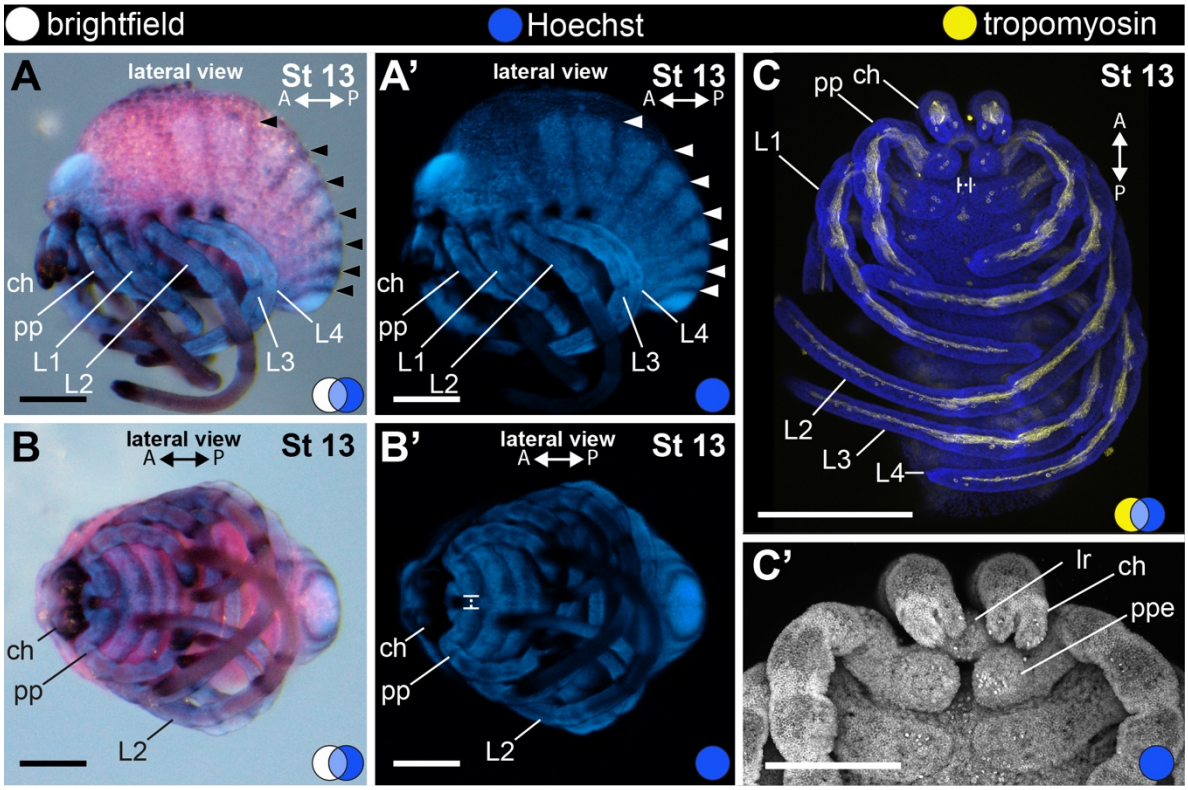


Figure 10: Stage 13 (**A–C′**). In situ hybridization for the gene *Popi-engrailed* (**A–B′**) and confocal micrographs of fluorescent antibody staining against tropomyosin (**C, C′**). **A–B′**: Seventh opisthosomal segment. **C**: Maximum projection of a fat-mounted embryo. **C′**: Detail of the chelicera and pedipalp (ventral view). A and P: anterior–posterior axis; arrowhead: stripe of engrailed expression marking the posterior border of opisthosomal segment; blue circle: Hoechst (blue, gray); white circle: bright field; yellow circle: tropomyosin (yellow); ch: cheliceral segment; lr: labrum; L1–4: leg 1–4 segments; pp: pedipalpal segment; ppe: pedipalpal endite. Scale bars **A–B′**: 100 µm. **C**: 250 µm. **C′**: 125 µm.

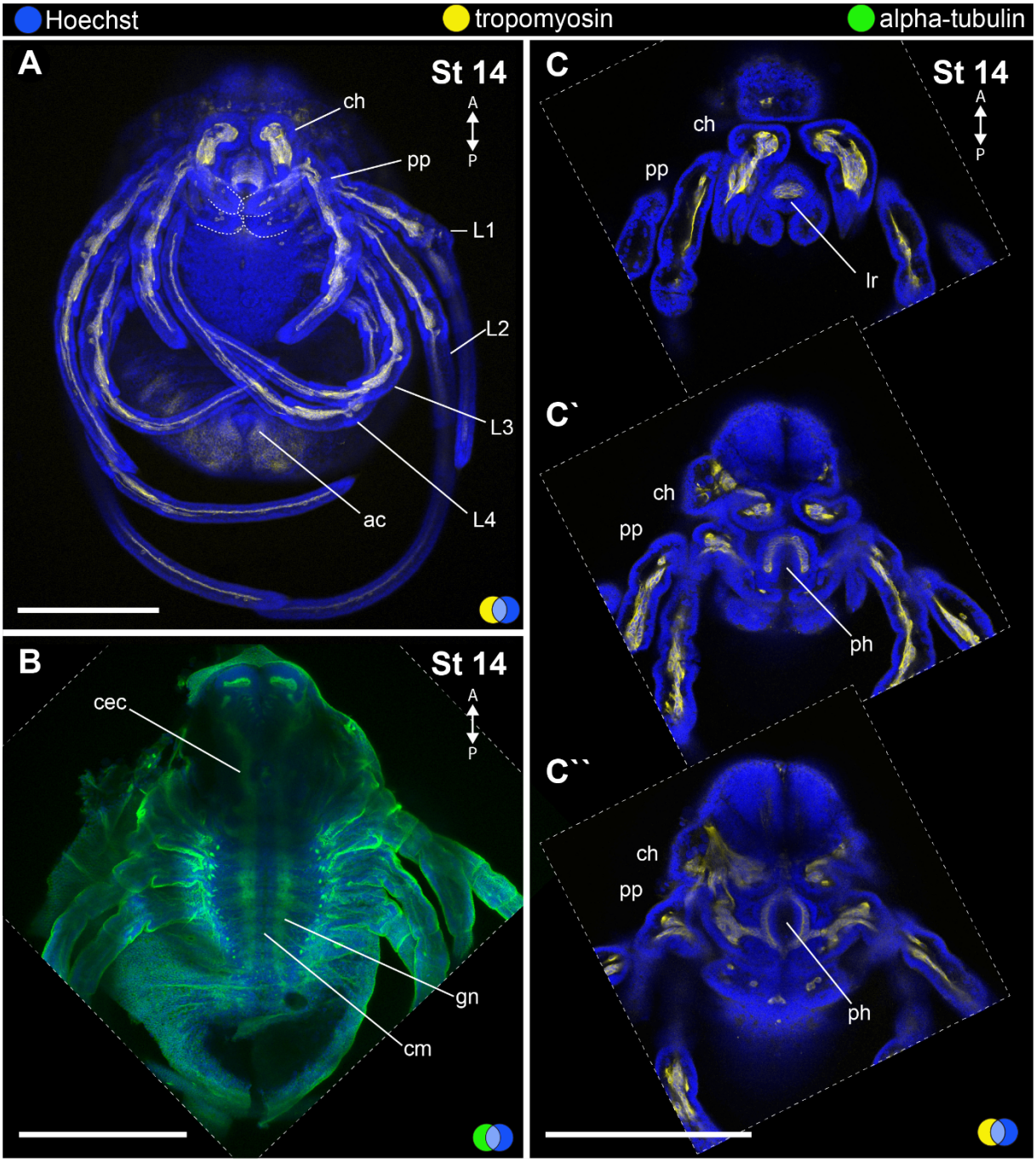


Figure 11: Stage 14 (**A–C″**). Confocal micrographs of fluorescent antibody staining against tropomyosin (**A, C–C″**) and alpha-tubulin (**B).** **A**: Maximum projection of fat-mounted embryo stained for tropomyosin and Hoechst (ventral view). Dotted lines mark the coxae of pedipalps and L1 legs. **B**: Maximum projection of mid-optical section of fat-mounted embryo stained for alpha-tubulin and Hoechst (ventral view). Anterior appendages have been dissected for clarity. **C–C″:** Optical sections of the head region of the same preparation in **A**. Note the tubular pharynx musculature. A and P: anterior–posterior axis; blue circle: Hoechst (blue); yellow circle: tropomyosin (yellow); green circle: tropomyosin (green); ca: corona analis. ch: chelicera segment; cec: circumesophageal connective; cm: commissure; gn: ganglium; lr: labrum; L1–4: leg 1–4 segments; ph: pharynx; pp: pedipalp. Scale bars: 250 µm. **B** reprinted with permission from (Sharma, 2018).

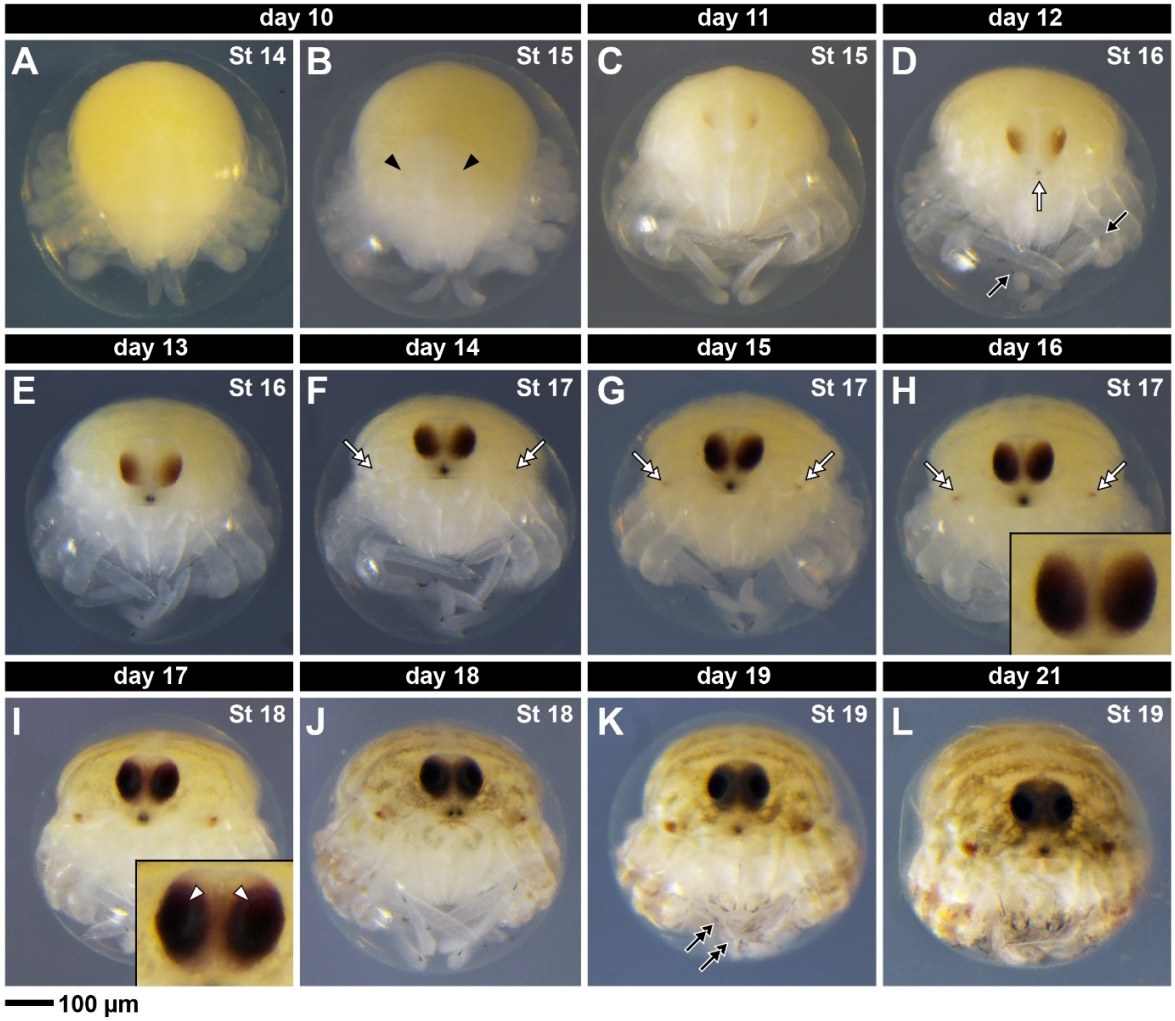


Figure 12: Stage 14 (**A**), stage 15 **(B–C**), stage 16 (**D–E**), stage 17 (**F–H**), stage 18 (**I–J**), and stage 19 (**K–L**). Frontal view of live embryos from the same clutch imaged under 1×PBS. Black arrow: tarsal claws; black double-arrow: setae; black arrowhead: eye pigment; white arrow: egg tooth; white double-arrow: ozopore; white arrowhead: eye lens. Scale bar: 100 µm.

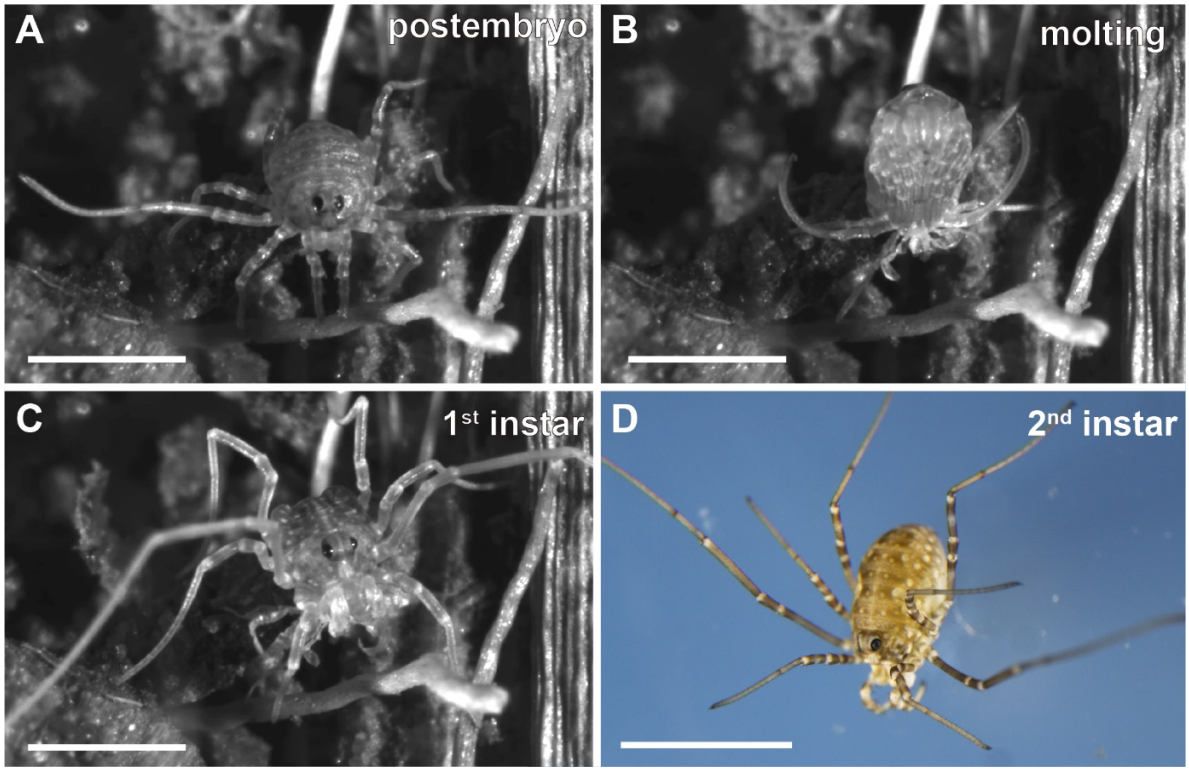


Figure 13: Postembryonic stages of *Phalangium opilio*. **A–C**: frames of a time-lapse imaging of postembryo moulting into first instar. **A**: postembryo. **B**: undergoing molting. **C**: 1st instar. **D**: 2nd instar. Scale bars: **A–D**: 1 mm.

Table 1. Duration of post embryonic instars, measured from data of a single clutch.

Stage	Time after hatching
Postembryo	day 0; n ≥ 3
1st instar	day 0 (2 h); n ≥ 3
2nd instar	day 7; n ≥ 3
3rd instar	day 13; n ≥ 3
4th instar	day 23; n = 2
5th instar	day 34; n = 2
6th instar	day 43; n = 2
Adult	#1(male): day 48; #2(female): day 55

Chapter 2

The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages

Guilherme Gainett[1†], Vanessa L. González[2†], Jesús A. Ballesteros[1], Emily V. W. Setton[1], Caitlin M. Baker[1], Leonardo Barolo Gargiulo[1], Carlos E. Santibáñez-López[3], Jonathan A. Coddington[2] and Prashant P. Sharma[1]

[1] Department of Integrative Biology, University of Wisconsin-Madison, Madison, 53706 WI, USA
[2] Global Genome Initiative, Smithsonian Institution, National Museum of Natural History, 10th and Constitution, NW, Washington, DC 20560-0105, USA
[3] Department of Biological and Environmental Sciences, Western Connecticut State University, 181 White St, Danbury, CT 06810, USA

Published August 2021 in Proceedings of the Royal Society B

Volume 288, Issue 1956, doi.org/10.1098/rspb.2021.1168

Abstract

Chelicerate arthropods exhibit dynamic genome evolution, with ancient whole-genome duplication (WGD) events affecting several orders. Yet, genomes remain unavailable for a number of poorly studied orders, such as Opiliones (daddy-long-legs), which has hindered comparative study. We assembled the first harvestman draft genome for the species *Phalangium opilio*, which bears elongate, prehensile appendages, made possible by numerous distal articles called tarsomeres. Here, we show that the genome of *P. opilio* exhibits a single Hox cluster and no evidence of WGD. To investigate the developmental genetic basis for the quintessential trait of this group—the elongate legs—we interrogated the function of the Hox genes *Deformed* (*Dfd*) and *Sex combs reduced* (*Scr*), and a homologue of *Epidermal growth factor receptor* (*Egfr*). Knockdown of *Dfd* incurred homeotic transformation of two pairs of

legs into pedipalps, with dramatic shortening of leg segments in the longest leg pair, whereas homeosis in L3 is only achieved upon double *Dfd* + *Scr* knockdown. Knockdown of *Egfr* incurred shortened appendages and the loss of tarsomeres. The similarity of *Egfr* loss-of-function phenotypic spectra in insects and this arachnid suggest that repeated cooption of EGFR signaling underlies the independent gains of supernumerary tarsomeres across the arthropod tree of life.

Introduction

The advent of genomic resources has revealed complex dynamics in the evolution of chelicerate genomes. A group of six terrestrial orders (Arachnopulmonata), which includes spiders, scorpions, and pseudoscorpions, exhibit an ancient shared whole-genome duplication (WGD), as evidenced by the architecture of Hox clusters, analyses of synteny, patterns of microRNA enrichment, gene expression patterns and gene tree topologies (Leite et al., 2016; Leite et al., 2018; Nolan et al., 2020; Ontano et al., 2021; Schwager et al., 2017; Sharma et al., 2014) (figure 1*a*). Separately, genomes of all four living Xiphosura (horseshoe crabs) suggest a lineage-specific twofold genome duplication in this order, with one of these duplications occurring relatively recently (Kenny et al., 2015; Shingate et al., 2020a; Shingate et al., 2020b). While genomes of Acariformes and Parasitiformes (mites and ticks) suggest that these two orders were not included in the genome duplication events, they often deviate from typical arthropod datasets. As examples, the acariform mite *Tetranychus urticae* exhibits extreme genome compaction (90 Mb), in tandem with the loss of many transcription factors, which has been linked to miniaturization (Grbić et al., 2011). Similarly, the genome of the parasitiform mite *Galendromus occidentalis* exhibits an atomized Hox cluster, degradation of synteny and high rates of intron gain and loss (Hoy et al., 2016).

One group that may facilitate comparative genomics of Chelicerata is the arachnid order Opiliones (harvestmen) (figure 1*b*). In phylogenomic datasets, Opiliones exhibit lower evolutionary rates than Parasitiformes or Acariformes, and their placement outside of arachnopulmonates makes this group phylogenetically significant (Ballesteros and Sharma, 2019; Lozano-Fernandez et al., 2019). Developmental transcriptomes of the emerging model species *Phalangium opilio* have suggested that harvestmen do not exhibit systemic genome duplication, as evidenced by the absence of paralogy across the homeobox gene family (Leite et al., 2018; Sharma et al., 2012) and gene expression patterns of genes with known paralogues

in arachnopulmonates (Leite et al., 2018; Nolan et al., 2020). As a result, *P. opilio* has proven useful for the study of chelicerate developmental biology. However, the establishment of a genome for this species is a key prerequisite to validating the assumption of an unduplicated genome in this order, as well as further advancing this model system.

Beyond its use in polarizing developmental traits on phylogenetic trees, Opiliones also exhibit a suite of unique characteristics that are not found in other arthropod models. The most salient of these are the elongate walking legs in some groups (e.g. Phalangioidea, commonly termed 'daddy-long-legs'). Beyond the hypertrophied growth of certain leg segments, many daddy-long-legs exhibit subdivision of the tarsus into pseudosegments called tarsomeres, with over 100 tarsomeres in some species (figure 1*c*). Tarsomeres have evolved dynamically across the arthropod tree of life, with gains in tarsomeres across insects (Kojima, 2017), scutigeromorph centipedes (Kenning et al., 2017) and several arachnid orders. However, the tarsomeres of daddy-long-legs are sufficiently numerous that they confer prehensility to the distal leg, which is used for climbing, courtship and male–male combat (figure 1*b*). The largest number of tarsomeres in Phalangioidea typically occurs on the antenniform second leg pair, which serves as a sensory appendage (Willemart et al., 2009) (figure 1*c*).

These aspects of harvestman biology position them as an opportune group for comparative study, both from the perspective of gene evolution before and after WGD, as well as understanding the genetic basis for morphological convergence (e.g. leg elongation; supernumerary tarsomeres). However, no genomes are available for any Opiliones. Moreover, the developmental genetic basis for arthropod leg elongation and tarsomere patterning is unknown outside of insects. To test the assumption that Opiliones exhibit an unduplicated genome, we generated a draft genome for *P. opilio* and leveraged this resource to investigate the genetic basis for leg patterning in this iconic arachnid group.

Material and methods

For brevity, detailed procedures, protocols and bioinformatic commands for the following operations are provided in the electronic supplementary material.

(a) Animal husbandry

For genome sequencing, founder population specimens of *P. opilio* were collected in Madison, WI, USA (43.074628, −89.403904), and a colony was maintained as previously described (Sharma et al., 2012) (electronic supplementary material, methods).

(b) RNA sequencing

Total RNA was extracted from *ca* 250 µl of *P. opilio* embryos spanning an array of stages, reared from females captured in Weston, MA, USA. RNA extraction was performed using TRIzol reagent (ThermoFisher), following the manufacturer's protocol. mRNA purification, library construction and 2 × 150 bp sequencing on an Illumina HiSeq 2500 platform follow our previous procedures (Sharma et al., 2014). The resulting 79 472 462 paired-end reads (NCBI PRJNA690950) were combined with an older library (16 225 145 paired-end reads sequenced on an Illumina GA II; NBCI PRJNA236471) for annotation of protein-coding regions.

(c) Genome sequencing, assembly and annotation

Full-sibling inbred lines were established at a colony in Madison, WI, USA. Total genomic DNA was isolated from two specimens (fourth-generation male and sixth-generation female). Long-read sequencing was performed on a PacBio Sequel platform (Pacific Biosciences) using v. 2.1 chemistry. The single-molecule real-time (SMRT) Cells were sequenced on 16 cells with 360 min movie lengths. Short-read sequencing was performed on

an Illumina HiSeq 2500 with a 350 bp insert size. Long reads were assembled using Canu v. 1.7 (Koren et al., 2017). Contigs were processed using two rounds of scaffolding with SSPACE-LongRead v. 1.1 (Boetzer and Pirovano, 2014), followed by gap filling with PBJelly v. 15.8 (English et al., 2012) and further polishing with Pilon v. 1.23 (Walker et al., 2014). Haplotypic duplicates were identified and removed with Purge_dups v. 1.2.3 (Guan et al., 2020). Prior to annotation, a custom repeat library was constructed using RepeatModeler open-1.0.11 (Smit et al., 2015) Identified repeats were masked with RepeatMasker open-4.0.6. (Smit et al., 2015). For annotation, gene predictions generated with BRAKER2 v. 2.1.5 (Hoff et al., 2019) used previously generated RNA-Seq reads from developmental transcriptomes (NCBI PRJNA236471; PRJNA690950; (Ballesteros and Sharma, 2019)). A genome browser was generated using MakeHub (Hoff, 2019).

(d) Orthology inference, phylogenetic analysis and discovery of microRNAs

Homologues of *Egfr* were identified with tBLASTn (Altschul et al., 1990), using query protein sequences of arthropod species for which *Egfr* expression has been previously studied. Sequence accession data are provided in the electronic supplementary material, table S1 and methods). An initial BLAST search for miRNA families in the genome of *P. opilio* and seven other chelicerates (electronic supplementary material, table S2) used as queries the miRNAs previously reported from the spider *Parasteatoda tepidariorum*, the tick *Ixodes scapularis* and the mite *T. urticae* (Leite et al., 2016; Ontano et al., 2021). To recover unique putative harvestmen miRNAs, we conducted a second search using miRNA families not known in spiders, ticks or mites, but which were shared by at least three mandibulate outgroups.

(e) Cloning, *in situ* hybridization and double-stranded RNA microinjection

Cloning of gene fragments, *in situ* hybridization and embryonic microinjections with dsRNA followed our previous procedures (Sharma et al., 2012; Sharma et al., 2013) (electronic

supplementary material, tables S3 and S4). Protocols for fluorescent gene expression assays using hybridization chain reaction (HCR) followed Bruce *et al.* (Bruce et al., 2021). For harvestman RNAi experiments, approximately two-thirds of each egg clutch was injected with dsRNA and the remaining third with water (negative control). Details of the phenotype scoring strategy are provided in the electronic supplementary material.

Results

(a) *Phalangium opilio* draft genome assembly

The draft assembly of the *P. opilio* genome comprises 580.4 Mbp (37.5% GC content) in 5137 scaffolds (N50: 211 089) and 8349 contigs (N50: 127 429; electronic supplementary material, figure S1 and table S5). The predicted genome repetitiveness is 54.4% and estimated heterozygosity is 1.24%. The number of predicted genes after filtering steps is 20 315, which was further refined with a 98% similarity threshold and manual curation to a final gene set of 18 036. This is comparable to predicted gene sets for the tick *I. scapularis* (20 486) and the mite *T. urticae* (18 414) (Grbić et al., 2011; Gulia-Nuss et al., 2016). An assessment using the arthropod set of benchmarking universal single-copy orthologues (BUSCOs) (Waterhouse et al., 2017) indicates 95.1% completeness (electronic supplementary material, figure S1 and table S5). Contamination assessment based on sequence coverage and GC content supports a relatively contamination-free assembly. The detailed description of the genome is provided in the electronic supplementary material, figures S1–S3.

(b) The genome of *Phalangium opilio* reveals the absence of Arachnopulmonata-specific whole-genome duplications

To assess whether *P. opilio* exhibits ancient WGD, we first examined the architecture

of Hox clusters in this species. We discovered one 480 kb scaffold that bore six Hox genes, with the remaining four Hox genes occurring on individual scaffolds (figure 2*a,b*). In addition to the small size, these scaffolds contained very few or no adjacent genes, suggesting that the position of these four Hox genes outside of the main cluster is an artefact of fragmentary assembly. On the larger 480 kb scaffold, microRNAs *mir-10* and *iab-4* were located adjacent to *Dfd* and *abdA*, respectively, which reflect conserved positions with respect to other arthropods (Pace et al., 2016) (figure 2*a*). The complete peptide sequences of all 10 Hox genes corresponded to previous partial sequences predicted from developmental transcriptomes (Sharma et al., 2012).

We separately examined the genome for evidence of duplicates in genes with known arachnopulmonate-specific paralogues and spatio-temporal subdivisions of expression patterns, focusing on four leg patterning genes (*dachshund*, *homothorax*, *extradenticle* and *spineless* (Nolan et al., 2020; Schwager et al., 2017; Setton et al., 2017)) and three retinal determination network genes (*sine oculis*, *Optix* and *orthodenticle* (Gainett et al., 2020; Samadi et al., 2015; Schomburg et al., 2015). These genes all occurred as single-copy in the harvestman genome (electronic supplementary material, table S6).

We next examined the distribution of families of microRNAs (miRNAs), noncoding RNAs with important regulatory roles in animals. miRNAs have been shown to exhibit the signature of genome duplication in both Arachnopulmonata and Xiphosura (Leite et al., 2016; Ontano et al., 2021). Thirty conserved miRNA families were identified in the *P. opilio* genome (figure 2*c*). Among them, only families mir-2, mir-29, mir-87 and mir-263 had two homologues in Opiliones (figure 2*c*). These microRNAs, with the exception of mir-29, are also duplicated in most other chelicerates and outgroup arthropods, (electronic supplementary material, table S2), suggesting the origin of paralogues at the arthropod common ancestor (figure 2*c*). The presence of duplicated mir-29 in harvestmen, horseshoe crabs and a subset of

Arachnopulmonata suggests separate independent duplication events in these lineages, although this parsimonious inference is contingent upon the resolution of the position of these groups in arachnid phylogeny. In sum, we found no evidence of miRNA duplications in the harvestman genome that were exclusively shared either with Arachnopulmonata or Xiphosura.

(c) *Deformed* and *Sex combs reduced* are necessary for leg fate specification in *Phalangium opilio*

Deformed homologues of arachnids are typically expressed in the four leg-bearing segments (L1–L4), whereas *Sex combs reduced* is expressed from L2 segment onwards (figure 3*a*; electronic supplementary material, figure S4) (Schwager et al., 2017; Sharma et al., 2012; Telford and Thomas, 1998). In the spider *P. tepidariorum* (a member of Arachnopulmonata), the two *Dfd* paralogues have divergent expression patterns; only *Ptep-DfdA* is expressed in the legs and expression levels are uniform across L1–L4 (Schwager et al., 2017). Knockdown of *Ptep-DfdA* results in homeotic L1-to-pedipalp transformation (Pechmann et al., 2015). No functional data exist for any single-copy homologue of *Dfd* in Arachnida, and no functional data exist for *Scr* in Arachnida altogether. Intriguingly, *Po-Dfd* is also expressed in L1–L4, but much more strongly in L2 (the longest leg pair) than the other three leg pairs, particularly during leg elongation (Sharma et al., 2012). *Scr* is expressed in L3 and L4, but its expression is much stronger in L3 (Sharma et al., 2012).

We first investigated the function of *Dfd* (*Po-Dfd*) through embryonic RNAi, via microinjection of double-stranded RNA (dsRNA) (electronic supplementary material, figure S5). Upon completion of embryogenesis, 33% (n = 84/248) of RNAi hatchlings exhibited leg-to-pedipalp homeosis affecting L1 and L2 (figure 3*b*,*c*; electronic supplementary material, figure S5). Complete leg-to-pedipalp homeotic transformation (figure 3*c*,*j*), only observed in L1, was evidenced by the loss of the metatarsus (segment specific to arachnid leg; absent in

pedipalp), the presence of pedipalp-specific setal spurs, the loss of the leg-specific tooth of the distal claw (figure 3*e,f,i,j*) and the undivided tarsus. Partial transformation was evidenced by a defective or absent metatarsus (figure 3*g–k*). Strongly affected L2 also exhibited a fusion of tibia and patella segments (figure 3*k*). Importantly, L3 and L4 were unaffected (figure 3*h*,*l*). The majority (n = 73/84) of *Po-Dfd* knockdown phenotypes exhibited mosaicism, with one side of the body more strongly affected (bilateral mosaics), so we further classified homeotic individuals in mosaic classes. This tabulation considered whether a given half presented no homeosis (wild-type), homeosis in both L1 and L2, homeosis in L1 only, or in L2 only (electronic supplementary material, figure S5). This analysis revealed that the most frequent homeotic conditions were having both legs affected (88/168 halves), followed by L1 only (30/168 halves) and L2 only (2/168 halves) (figure 3*q*; electronic supplementary material, figure S5). In halves with both legs affected, L1 was always more strongly transformed than L2. Notably, knockdown of *Po-Dfd* resulted in dramatic shortening of the transformed appendages, and particularly for transformed L2 (compare figure 3*g*,*k*). Reduced *Po-Dfd* expression was observed overall in embryos injected with *Po-Dfd* dsRNA and correlated with the side presenting homeotic transformation in mosaic embryos (electronic supplementary material, figure S6).

Next, to investigate a possible role of *Scr* in the identity of L3 and L4, we performed RNAi against *Po-Scr*. Despite verifiable decreased expression (electronic supplementary material, figure S6), we detected no phenotypic effects on dsRNA-injected embryos (five clutches, n = 540 embryos), suggesting that *Scr* may exhibit functional redundancy in arachnids. We therefore performed a double knockdown of *Po-Dfd* and *Po-Scr*. Mortality was high (153/176), but double RNAi resulted in partial leg-to-pedipalp transformation affecting L1, L2 and L3 (n = 8/176) (figure 3*d*, *m–q*; electronic supplementary material, figure S5).

Taken together, these results suggest that *Po-Dfd* is necessary for conferring leg

identity of the L1 and L2 segments, and that both *Dfd* and *Scr* are necessary for the identity of the L3 segment.

(d) Epidermal growth factor receptor is necessary for distal leg patterning in the daddy-long-legs

A notable component of leg morphology in daddy-long-legs is the repeated subdivision of the tarsus, the distalmost leg segment. In insects, EGFR signaling is involved in tarsal fate specification (Campbell, 2002; Galindo et al., 2002; Galindo et al., 2005), but it is unknown if this signaling pathway is necessary for leg patterning in chelicerates (figure 4*a*). Upon surveying the *P. opilio* genome, we discovered two *Egfr* paralogues in the harvestman (electronic supplementary material, figure S7). One of these, *Po-EgfrB*, lacks the transmembrane and intracellular domains seen in other *Egfr* homologues (electronic supplementary material, figure S8). A 3′ UTR for *Po-EgfrB* was assembled in both embryonic transcriptomes and corroborated by the genome assembly, disfavouring fragmentary assembly as a possible explanation for missing domains. We therefore focused on *Po-EgfrA*, the paralogue containing all known *Egfr* functional domains.

In early limb bud stage, *Po-EgfrA* is expressed in a strong circular domain around the stomodeum, in a strong ring at the base of each appendage and in a weak stripe along the ventral midline (figure 4*b*; electronic supplementary material, figure S9). In later stages, additional expression domains occur in the developing eye field, in a strong domain in the medial bridge of the developing brain, and in rings at the boundaries of the segments (podomeres) of the developing appendages (figure 4*c*,*d*; electronic supplementary material, figure S9). To investigate EGFR signaling further, we surveyed the expression of a homologue of *pointed*, an ETS transcription factor that acts as an EGFR signal effector (Brunner et al., 1994). In early limb bud stages, the single-copy *Po-pnt* is expressed in the ventral ectoderm

and the distal tip of the appendages (figure 4*e*; electronic supplementary material figure S9). Similar to the beetle *T. castaneum* (Grossmann and Prpic, 2012), in later stages, *Po-pnt* is also expressed in the head lobes, and groups of cells in the appendages, particularly in the distal region, forming rings (figure 4*f*,*g*; electronic supplementary material, figure S9).

RNAi against *EgfrA* resulted in 39.6% (n = 36/91) of hatchlings exhibiting segmentation and appendage defects (figure 4*h*–*t*; electronic supplementary material, figure S10); all affected embryos were bilateral mosaics (figure 4*i*,*j*) and defects correlated with reduced *Po-EgfrA* expression in embryos (electronic supplementary material, figure S11). *Po-EgfrA* dsRNA-injected hatchlings showed defects of antero-posterior (body) segmentation, with dorsal tissue fusion on the side of the body affected (n = 29/36), correlating with a characteristic curved shape of the body of mosaic individuals (figure 4*i*). Defects of the eyes (n = 25/36) ranged from a small reduction in size to complete absence (figure 4*j*).

The appendages exhibited an array of defects in terminal structures (figure 4*k*–*t*). The chelicera showed the reduction of the fixed finger, movable finger or both (n = 11/36) (figure 4*o*); the pedipalp showed a reduced claw (n = 19/36) (figure 4*p*); and, in the case of all legs, the claw and tarsomeres were reduced (n = 26/36) (figure 4*q*,*r*). Most notably, a subset of weakly affected individuals (n = 10) showed a condition in which the distalmost tarsomeres were fused and the claw was missing, or just the claw was missing (figure 4*t*). Segmental fusions in limbs occurred in a subset of hatchlings (n = 14/36) (figure 4*q*). The proximal segment of the appendages (coxa) showed defects ranging from reduction in size to complete proximal fusion of adjacent appendage coxae (n = 34/36) (figure 4*s*), which correlated with the strong ring of expression at the base of all appendages.

These results are broadly consistent with expression and functional data available for *Egfr* homologues in insect models (Grossmann and Prpic, 2012; Halfar et al., 2001; Nakamura et al., 2008; Refki and Khila, 2015). The loss-of-function phenotypic spectrum in *P. opilio*

suggests that *EgfrA* may underpin both leg elongation and tarsomere morphogenesis in daddy-long-legs.

Discussion

(a) Genomics and Hox-logic in a phylogenetically significant arachnid model

Spatio-temporal subdivision of expression domains of duplicated developmental patterning genes are systemic in Arachnopulmonata, as established by gene expression surveys in model spiders, scorpions and whip spiders (Gainett and Sharma, 2020; Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2014). This phenomenon makes arachnopulmonates an ideal taxon for investigating the role of gene duplication in generating body plan disparity. Identifying subfunctionalization or neofunctionalization of duplicates requires a clear inference of the ancestral single-copy homologue's expression pattern and function. In this regard, *P. opilio* has played a central role in polarizing developmental phenomena, given its phylogenetic position, low evolutionary rate and tractability in the laboratory (Baudouin-Gonzalez et al., 2021; Leite et al., 2018; Nolan et al., 2020).

Nevertheless, the assumed unduplicated condition of the Opiliones genome has not been rigorously tested. The efficiency of developmental transcriptomes in discovering paralogy is a function of sequencing depth and sampling strategy, and thus transcriptomes frequently fail to capture paralogues (Leite et al., 2018; Ontano et al., 2021; Sharma et al., 2014). As a first step in validating the use of *P. opilio* as an effective outgroup to Arachnopulmonata, we assembled and interrogated the genome of *P. opilio*, which revealed no evidence of systemic genome duplication events previously reported for arachnopulmonates or horseshoe crabs. Together with the genome architecture of model systems like *I. scapularis* and *T. urticae*, the condition of the harvestman genome strongly

supports the inference that an unduplicated genome is the ancestral condition for arachnids.

Having established that Hox genes of *P. opilio* are *bona fide* single-copy, we targeted the first functional datapoints for single-copy Hox genes in any arachnid species, focusing on genes that pattern walking leg identity. We were able to show that the knockdown of *Po-Dfd* affects the identity of legs 1 and 2. This discovery is significant for two reasons. First, two copies of *Dfd* occur in spiders and these exhibit subdivision of expression pattern, with *Ptep-DfdA* being expressed throughout leg tissues and ventral ectoderm, and *Ptep-DfdB* mostly restricted to the ventral ectoderm (Schwager et al., 2017). It has been shown that the spider paralogue *Ptep-DfdA* is necessary for repressing pedipalp identity on a single-body segment (L1); there is no effect on L2 (Pechmann et al., 2015). Compared to the *Po-Dfd* RNAi phenotype, these data are congruent with the hypothesis that arachnopulmonate *Dfd* paralogues have undergone subfunctionalization. Further functional studies of *DfdB* paralogues in arachnopulmonates are necessary to test this scenario, specifically targeting *DfdB* (alone and also through double knockdown with *DfdA*). Second, segments affected by *Dfd* knockdown in harvestman (L1 and L2) are positionally homologous to those affected by *Dfd* knockdown in pancrustaceans (mandibular and maxillary segments (Brown et al., 2000; Hughes CL and Kaufman TC, 2002; Martin et al., 2016)). These results bring further support for the notion that the establishment of some Hox anterior boundaries predates the evolution of tagmata, with further substantiation from Hox anterior boundaries in Onychophora (Janssen et al., 2014).

Furthermore, we found that the knockdown of *Po-Scr* alone has no discernible phenotype, whereas the double knockdown of *Po-Dfd* and *Po-Scr* resulted in homeotic transformation of L1–L3 into pedipalps. These data suggest functional redundancy of *Scr* in L3, paralleling the dynamics of *Ubx* and *abdA* in insects; the knockdown of *abdA* alone has a limited effect on homeotic abdominal-to-thoracic segment transformations, whereas the

double knockdown of both these genes results in more complete transformations of abdominal segments into thoracic identities (Angelini et al., 2005). Similar functional redundancy has been shown for *Antp-1* and *Ubx-1* in the spider *P. tepidariorum* (Khadjeh et al., 2012).

As these experiments show, *P. opilio* has the potential to serve as an informative outgroup to Arachnopulmonata for the study of paralogue divergence after duplication. In addition to newly generated genomic resources, the effectiveness of single and double RNAi in this system makes *P. opilio* an opportune point of comparison for future investigations of arachnid body plan evolution.

(b) A conserved role for epidermal growth factor receptor signalling in appendage patterning across Arthropoda

In the fruit fly *Drosophila melanogaster*, EGFR-Ras signalling is responsible for patterning the legs in two phases. First, EGFR signalling begins with a distal expression (central in the leg disc) of the EGFR ligands in the leg disc (Campbell, 2002; Galindo et al., 2002). The reduction of EGFR signalling results in progressively greater deletion of more distal leg structures, which is consistent with a distal-to-proximal requirement gradient of EGFR signalling by downstream tarsal patterning genes, and with a distal source of EGFR signalling (Campbell, 2002; Galindo et al., 2002; Kojima, 2017). A distal source of EGFR signalling is also in accordance with the distal-to-proximal requirement for EGFR signalling in the regenerating leg of the cricket *G. bimaculatus* (Nakamura et al., 2008). In *P. opilio*, 72% (n = 26/36) of the hatchlings resulting from RNAi against *Egfr* exhibited defects in the tarsus. These defects ranged from the absence of claws and distal tarsomere fusions, to complete failure to form all tarsal subdivisions. The spectrum of increasingly severe defects from distal-to-proximal observed upon *Egfr* knockdown in *P. opilio* suggests that daddy-long-leg tarsomeres are patterned by a gradient of EGFR signalling similar to *D. melanogaster* and *G.*

bimaculatus. This conclusion is further supported by the expression of the EGFR signalling effector *Po-pnt*, which we showed to occur on the tip of the developing appendages at the limb bud stage.

In a second phase, EGFR ligands are expressed as rings at the boundaries of embryonic tarsomeres (Galindo et al., 2005). The reduction or upregulation of EGFR signalling at this later stage in *D. melanogaster* results in defects in medial tarsomeres and in failure to correctly pattern the tarsal joints (Campbell, 2002; Galindo et al., 2002). However, in short germ insect models (e.g. beetle, water strider and cricket), *Egfr* is expressed as rings at the boundaries of all leg segments proximal to the tarsus, in contrast to rings restricted to the tarsus in the fruit fly (Grossmann and Prpic, 2012; Nakamura et al., 2008; Refki and Khila, 2015). RNAi-mediated knockdown in two insect models resulted in leg segment truncations proximal to the tarsus (Grossmann and Prpic, 2012; Refki and Khila, 2015). Our data in the arachnid *P. opilio* largely accord with the expression and functional results in short germ insects: 38% (n = 14/36) of *Po-EgfrA* RNAi phenotypes also exhibited leg segment fusions proximal to the tarsus. The signal effector *Po-pnt* is also expressed in a distal domain forming rings in later stages of appendage development, which accords with the second phase of expression in model insects (Galindo et al., 2005; Grossmann and Prpic, 2012). Together, these results suggest that the second role of EGFR signalling in leg segmentation may also be conserved in *P. opilio*.

Disentangling the effects of early versus late EGFR signaling phases in the phenotypes observed in *Po-EgfrA* RNAi could be further explored by disrupting EGFR signaling in later development, to surpass the early function in PD axis development. We anticipate that the genome of *P. opilio* will facilitate the development of more sophisticated tools for functional genetics, toward refining the understanding of how daddy-long-legs make their long legs.

Data accessibility

The data underlying this article are available in the article and in its online supplementary material. The genome assembly, EGFR tree and alignment, are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.ht76hdrds (Gainett et al., 2021a). Sequencing data are deposited in NCBI under accession numbers NCBI PRJNA690950 (RNA-seq), NCBI PRJNA647749 (genome), NCBI SRR12286133 (long reads), NCBI SRR12286133 (short reads). The genome browser is available at: http://phalangium-opilio.sigenomehub.org/. The data are provided in electronic supplementary material (Gainett et al., 2021b).

Authors' contributions

G.G.: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing-original draft, writing-review and editing; V.L.G.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, visualization, writing-original draft, writing-review and editing; J.B.: data curation, formal analysis, methodology, software, writing-review and editing; E.V.W.S.: data curation, methodology, writing-review and editing; C.M.B.: resources, writing-review and editing; L.B.G.: data curation, investigation, writing-review and editing; C.E.S.-L.: data curation, formal analysis, investigation, methodology, writing-review and editing; J.A.C.: funding acquisition, resources, supervision, writing-review and editing; P.S.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing-original draft, writing-review and editing

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests

We declare we have no competing interests.

Funding

This material is based on work supported by the Food and Drug Administration (J.A.C.), the Global Genome Initiative grant no. GGI-Exploratory-2016-047 (V.L.G.), and National Science Foundation grant nos. IOS-1552610 and IOS-2019141 (P.P.S.). G.G. was supported by a Wisconsin Alumni Research Foundation Fall Research Competition award.

Acknowledgements

RNA sequencing was performed at the Center for Systems Biology, Harvard University. Genomic laboratory work was conducted at the Smithsonian Laboratories of Analytical Biology (LAB). Specimen vouchering and long-term cryogenic curation was provided by the Smithsonian National Museum of Natural History. HCR was made possible by the Patel lab and the 2021 MBL embryology course. Microscopy was performed at the Newcomb Imaging Center, Department of Botany, University of Wisconsin-Madison. Computing was conducted by the Smithsonian Institution High Performance Cluster (https://doi.org/10.25572/SIHPC). Audrey R. Crawford and Calvin So assisted with harvestman rearing and RNAi experiments.

Footnotes

[†]Co-first authors.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5534425.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol* 215, 403–410.

Angelini, D. R., Liu, P. Z., Hughes, C. L. and Kaufman, T. C. (2005). Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). *Dev. Biol.* 287, 440–455.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440–22.

Baudouin-Gonzalez, L., Schoenauer, A., Harper, A., Blakeley, G., Seiter, M., Arif, S., Sumner-Rooney, L., Russell, S., Sharma, P. P. and McGregor, A. P. (2021). The evolution of Sox gene repertoires and regulation of segmentation in arachnids. *Mol. Biol. Evol*. 38, 3153–3169.

Boetzer, M. and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15, 211.

Brown, S., DeCamillis, M., Gonzalez-Charneco, K., Denell, M., Beeman, R., Nie, W. and Denell, R. (2000). Implications of the *Tribolium Deformed* mutant phenotype for the evolution of Hox gene function. *Proc. Natl Acad. Sci. USA* 97, 4510–4514.

Bruce, H. S., Jerz, G., Kelly, S. R., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D. A., Wolff, C. and Patel, N. H. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol V.1.

Brunner, D., Dücker, K., Oellers, N., Hafen, E., Scholzi, H. and Klämbt, C. (1994). The ETS domain protein Pointed-P2 is a target of MAP kinase in the Sevenless signal transduction pathway. *Nature* 370, 386–389.

Campbell, G. (2002). Distalization of the *Drosophila* leg by graded EGF-receptor activity. *Nature* 418, 781–785.

Chipman, A. D., Ferrier, D. E. K., Brena, C., Qu, J., Hughes, D. S. T., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F. C., et al. (2014). The first myriapod genome sequence reveals conservative arthropod gene content and genome organization in the centipede *Strigamia maritima*. *PLOS Biol.* 12, e1002005–24.

English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7, e47768.

Gainett, G. and Sharma, P. P. (2020). Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *Evodevo* 11, 18–18.

Gainett, G., Ballesteros, J. A., Kanzler, C. R., Zehms, J. T., Zern, J. M., Aharon, S., Gavish-Regev, E. and Sharma, P. P. (2020). Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genom.* 21, 811–17.

Gainett, G., González, V. L., Ballesteros, J. A., Setton, E. V. W., Baker, C. M., Gargiulo, L. B., López, C. E. S., Coddington, J. A. and Sharma, P. P. (2021a). The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages and chelicerate genome architecture. *Dryad, Dataset*.

Gainett, G., González, V. L., Ballesteros, J. A., Setton, E. V. W., Baker, C. M., Gargiulo, L. B., López, C. E. S., Coddington, J. A. and Sharma, P. P. (2021b). The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages and chelicerate genome architecture. *Figshare*.

Galindo, M. I., Bishop, S. A., Greig, S. and Couso, J. P. (2002). Leg patterning driven by

proximal-distal interactions and EGFR signaling. *Science (1979)* 297, 256–259.
Galindo, M. I., Bishop, S. A. and Couso, J. P. (2005). Dynamic EGFR-Ras signalling in *Drosophila* leg development. *Dev. Dyn.* 233, 1496–1508.
Grbić, M., Leeuwen, T. Van, Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.
Grossmann, D. and Prpic, N.-M. (2012). Egfr signaling regulates distal as well as medial fate in the embryonic leg of *Tribolium castaneum*. *Dev. Biol.* 370, 264–272.
Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y. and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36, 2896–2898.
Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., Sattelle, D. B., Fuente, J. D. La, Ribeiro, J. M., Megy, K., et al. (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* 7, 10507.
Halfar, K., Rommel, C., Stocker, H. and Hafen, E. (2001). Ras controls growth, survival and differentiation in the *Drosophila* eye by different thresholds of MAP kinase activity. *Development* 128, 1687–1696.
Hoff, K. J. (2019). MakeHub: fully automated generation of UCSC genome browser assembly hubs. *Genom. Proteom. Bioinform.* 17, 546–549.
Hoff, K. J., Lomsadze, A., Borodovsky, M. and Stanke, M. (2019). Whole-Genome annotation with BRAKER. *Methods Mol. Biol.* 1962, 65–95.
Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., Pomerantz, A. F., Simão, F. A., Thomas, J., Jiggins, F. M., et al. (2016). Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized Hox genes and superdynamic intron evolution. *Genome Biol. Evol.* 8, 1762–1775.
Hughes CL and Kaufman TC (2002). RNAi analysis of *Deformed, proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the Hemipteran head. *Development* 127, 3683–3694.
Janssen, R., Eriksson, B. J., Tait, N. N. and Budd, G. E. (2014). Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front. Zool.* 11, 22.
Kenning, M., Müller, C. H. G. and Sombke, A. (2017). The ultimate legs of Chilopoda (Myriapoda): a review on their morphological disparity and functional variability. *PeerJ* 5, e4023–36.
Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., et al. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116, 190–199.
Khadjeh, S., Turetzek, N., Pechmann, M., Schwager, E. E., Wimmer, E. A., Damen, W. G. M. and Prpic, N. M. (2012). Divergent role of the Hox gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl Acad. Sci. USA* 109, 4921–4926.
Kojima, T. (2017). Developmental mechanism of the tarsus in insect legs. *Curr. Opin. Insect. Sci.* 19, 36–42.
Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736.
Leite, D. J., Ninova, M., Hilbrant, M., Arif, S., Griffiths-Jones, S., Ronshaugen, M. and McGregor, A. P. (2016). Pervasive microRNA duplication in chelicerates: insights from the embryonic microRNA repertoire of the spider *Parasteatoda tepidariorum*.

Genome Biol. Evol. 8, 2133–2144.

Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N.-M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018). Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* 35, 2240–2253.

Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D. and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* 10, 2295–8.

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C. and Patel, N. H. (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Curr. Biol.* 26, 14–26.

Nakamura, T., Mito, T., Miyawaki, K., Ohuchi, H. and Noji, S. (2008). EGFR signaling is required for re-establishing the proximodistal axis during distal leg regeneration in the cricket *Gryllus bimaculatus* nymph. *Dev. Biol.* 319, 46–55.

Nolan, E. D., López, C. E. S. and Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Dev. Genes. Evol.* 230, 137–153.

Ontano, A. Z., Gainett, G., Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., et al. (2021). Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol. Biol. Evol.* 38, 2446–2467.

Pace, R. M., Grbić, M. and Nagy, L. M. (2016). Composition and genomic organization of arthropod Hox clusters. *Evodevo* 7, 1–11.

Pechmann, M., Schwager, E. E., Turetzek, N. and Prpic, N.-M. (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc R. Soc. B.* 282, 20151162.

Refki, P. N. and Khila, A. (2015). Key patterning genes contribute to leg elongation in water striders. *Evodevo* 6, 14.

Samadi, L., Schmid, A. and Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *Evodevo* 6, 16.

Schomburg, C., Turetzek, N., Schacht, M. I., Schneider, J., Kirfel, P., Prpic, N.-M. and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *Evodevo* 6, 15.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.

Setton, E. V. W., March, L. E., Nolan, E. D., Jones, T. E., Cho, H., Wheeler, W. C., Extavour, C. G. and Sharma, P. P. (2017). Expression and function of *spineless* orthologs correlate with distal deutocerebral appendage morphology across Arthropoda. *Dev. Biol.* 430, 224–236.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.

Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones).

Evol. Dev. 15, 228–242.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Wheeler, W. C. (2014). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661.
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H. and Venkatesh, B. (2020a). Chromosome-level genome assembly of the coastal horseshoe crab (*Tachypleus gigas*). *Mol. Ecol. Resour.* 20, 1748–1760.
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E. and Venkatesh, B. (2020b). Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.* 11, 2322.
Smit, A., Hubley, R. and Green, P. (2015). *RepeatMasker Open-4.0* 2013-2015.
Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* 95, 10671–10675.
Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963.
Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V and Zdobnov, E. M. (2017). BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548.
Willemart, R. H., Farine, J.-P. and Gnaspini, P. (2009). Sensory biology of Phalangida harvestmen (Arachnida, Opiliones): a review, with new morphological data on 18 species. *Acta. Zool.* 90, 209–227.

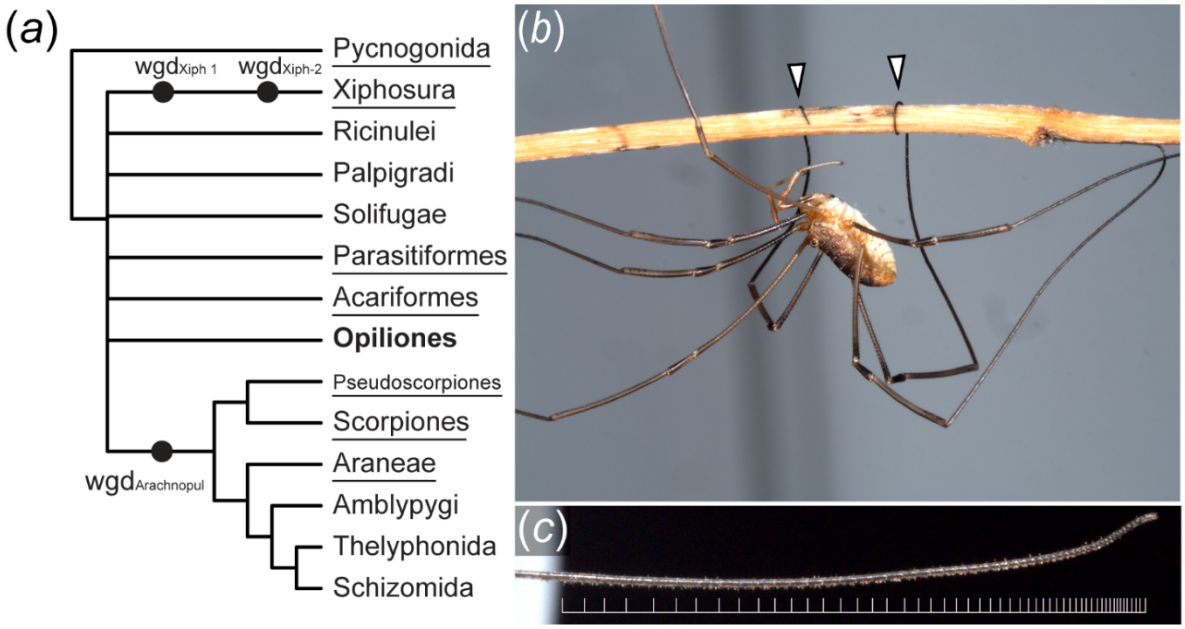


Figure 1: The significance of Opiliones in evolutionary developmental biology. (a) Consensus phylogeny of Chelicerata (based on Ontano et al., 2021) and inferred WGD events in Xiphosura and Arachnopulmonata. (b) Adult male *P. opilio* climbing on a twig using its prehensile tarsi. (c) Detail of the distal subdivisions (tarsomeres) of the leg 2 tarsus. The distal terminus is to the right. White bars mark tarsomere boundaries. Photographs: Caitlin M. Baker.

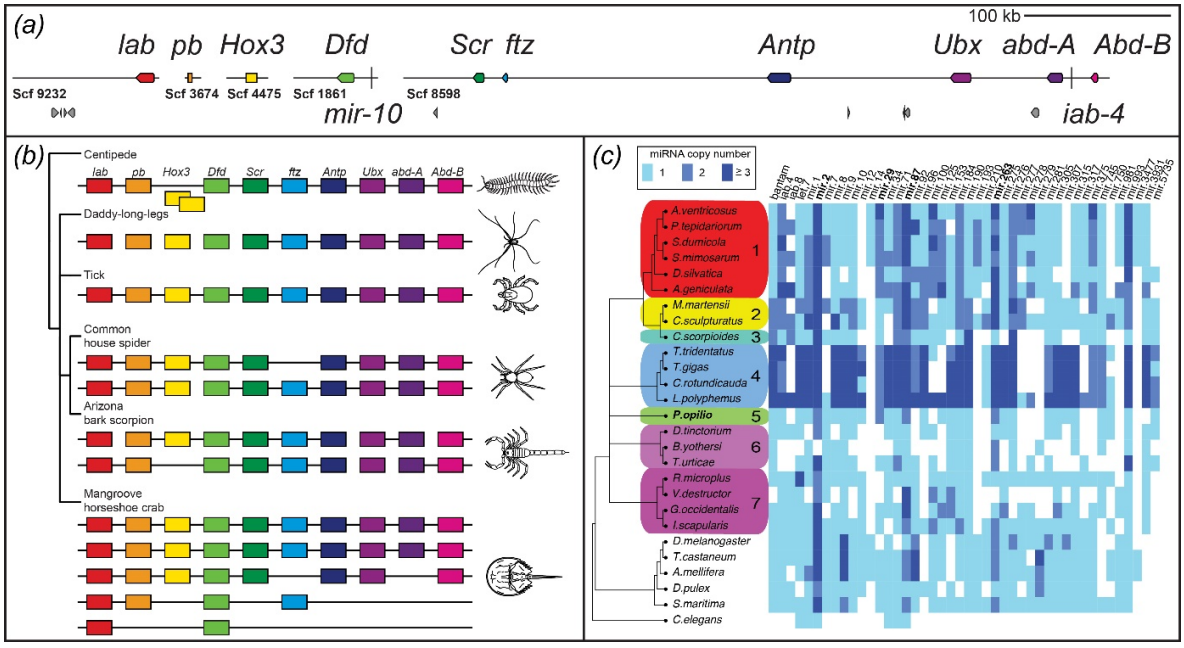


Figure 2: Hox genes and microRNAs support an unduplicated genome in the daddy-long-legs *P. opilio*. (a) Hox gene-containing scaffolds to scale. miRNAs mir-10 and iab-4 are represented by vertical bars. Hox genes are depicted in coloured boxes, and other predicted genes in grey boxes. (b) Hox clusters in selected arthropod genomes (after Chipman et al., 2014; Ontano et al., 2021). (c) Comparative analysis of miRNA families and orthologue copy numbers in *P. opilio* and other chelicerates supports retention of single copies of several families in harvestmen, in contrast to duplication found in arachnopulmonates. Columns correspond to individual miRNA families, with colours representing a different number of paralogues. miRNAs in bold are duplicated in *P. opilio* and most other chelicerates. 1: Araneae; 2: Scorpiones; 3: Pseudoscorpiones; 4: Xiphosura; 5: Opiliones; 6: Acariformes; 7: Parasitiformes.

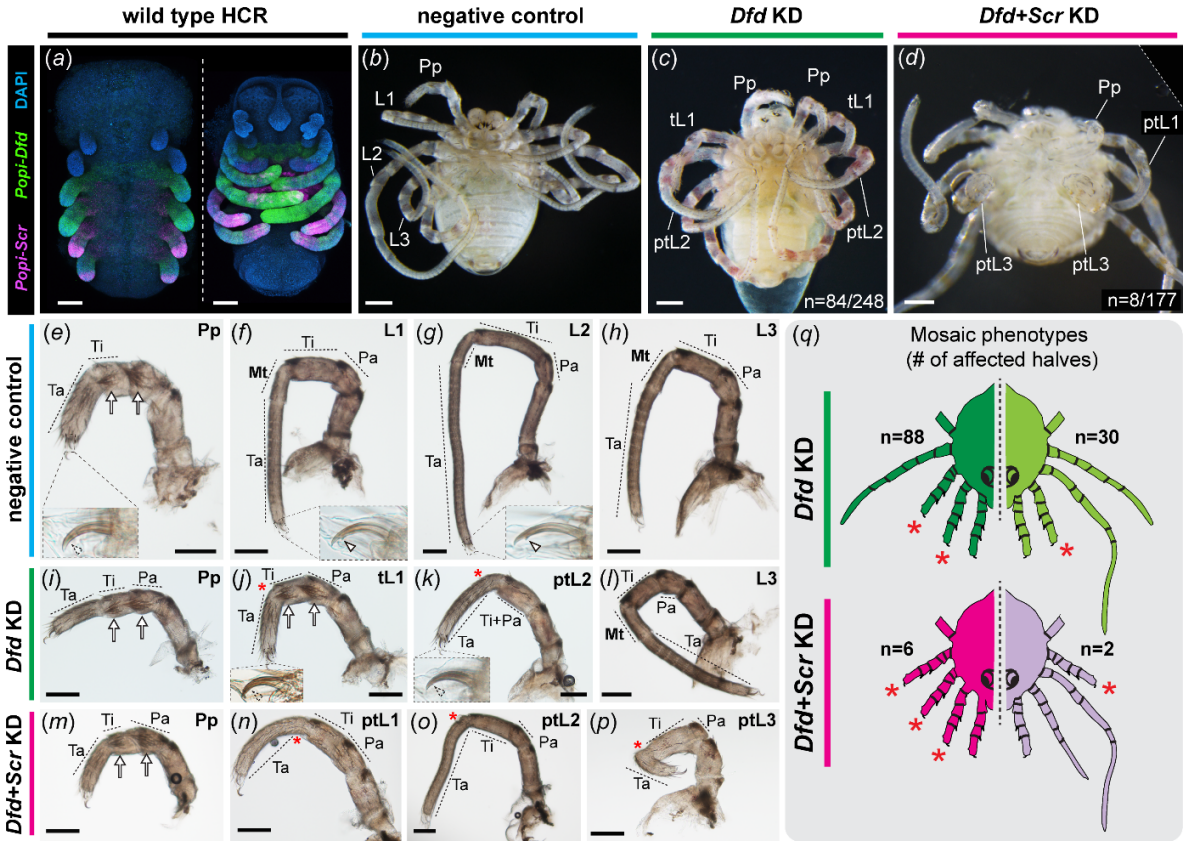


Figure 3: The single-copy orthologue of *Deformed (Dfd)* and *Sex combs reduced (Scr)* in *P. opilio* are necessary for the appendage identity of three body segments. (a) Wild-type HCR expression patterns of *Dfd* (green) and *Scr* (magenta) homologues in stage 11 and 15 embryos. (c) Wild-type (negative control) hatchling of *P. opilio*. (c) *P. opilio* hatchling from *Po-Dfd* RNAi treatment. (d) *P. opilio* hatchling from *Po-Dfd* + *Scr* RNAi treatment. (e–h) Appendage mounts of wild-type *P. opilio* hatchlings in lateral view. Insets: tarsal claws. (i–l) Appendage mounts of *Po-Dfd* RNAi hatchlings in lateral view. (m–p) Appendage mounts of *Po-Dfd* + *Scr* RNAi hatchlings in lateral view. (q) Schematic depiction of the mosaicism observed in hatchlings of *Po-Dfd* RNAi (*Dfd* KD) and *Po-Dfd* + *Scr* RNAi (*Dfd* + *Scr* KD) and distribution of phenotypic classes. Lighter colours indicate weaker penetrance. Only the two most frequent homeotic conditions are depicted (see electronic supplementary material, figure S5). Asterisk, reduced metatarsus; arrow, setal spurs; arrowhead, claw tooth; L1–L3: leg 1–3; tL1, fully transformed leg 1; ptL1–3: partially transformed legs 1–3; Mt, metatarsus; Pa, patella; Pp, pedipalp; Ta, tarsus; Ti; tibia. Scale bars: 100 µm.

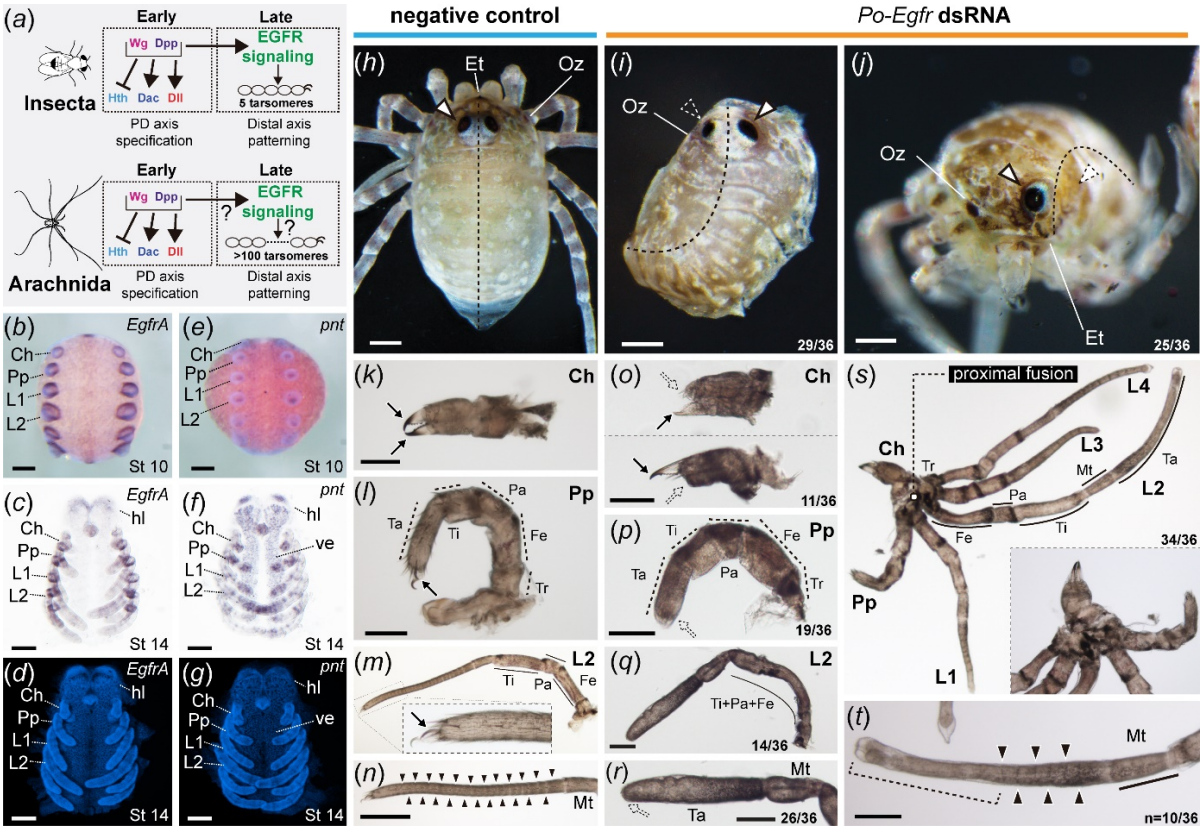


Figure 4: *Po-EgfrA* knockdown affects dorsal patterning, eyes and appendage formation. (a) Gene regulatory network specifying PD axis and distal appendage patterning in insects and arachnids. *Po-EgfrA* (b–d) and *Po-pnt* (e–g) in situ hybridization wild-type expression. (b,e) Whole-mount stage 10 embryos, merged brightfield and Hoechst nuclear staining, ventral view. (c,f) Flatmounts, stage 14 embryos, brightfield, ventral view. (d,g) Hoechst nuclear counter staining. (h) Negative control hatchling in dorsal view. (h,j) Hatchlings from *Po-EgfrA* dsRNA-injected treatment (mosaic, left side affected). (b) Hatchling in dorsal view. Note dorsal fusion on the left side of the body (n = 29/36). (c) Hatchling in frontal view, with the left eye absent. A subset of *Egfr* phenotypes showed eye reduction (25/36) (k–n) Appendage flat mounts of negative control hatchlings, in lateral view. (k) Chelicera. (l) Pedipalp. (m) L2. Inset: detail of the claw. (n) Tarsus of L2. (o–t) Appendage flat mounts of hatchlings of *Po-EgfrA* dsRNA-injected treatment, in lateral view. (o) Chelicerae with a reduced fixed finger (upper panel), movable finger (lower panel) or both (n = 11/36). (p) Pedipalps lacking claw (n = 19/36). (q) L2, exhibiting podomere fusions proximal to the tarsus (n = 14/36). (r) Distal end of L2, exhibiting claw and tarsomere reduction (n = 26/36). (s) Proximal fusion in adjacent appendages (Ch–L4) (n = 34/36). Inset: Detail of fused coxae. (t) Tarsus of leg 2 shown in (r). Weakly affected legs lacked claws and distal tarsal joints (brackets) but retained proximal joints (n = 10/36). Arrow, claw; outlined white arrowhead, eye; dotted white arrowhead, eye defect; solid black arrowhead, tarsomere joints; Ch, chelicera; Et, egg tooth; Fe, femur; hl, head lobe; L1–L4, legs 1–4; Mt, metatarsus; Oz, ozophore; Pa, patella; Pp, pedipalp; Ta, tarsus; Ti, tibia; Tr, trochanter; ve, ventral ectoderm. Scale bars: 100 µm.

Chapter 3

Dual functions of *labial* resolve the Hox logic of chelicerate head segments

Guilherme Gainett[1], Benjamin C. Klementz[1], Pola O. Blaszczyk[1], Heather S. Bruce[2], Nipam H. Patel[2,3], Prashant P. Sharma[1]*

[1]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA 53706
[2]Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA
[3]University of Chicago, Organismal Biology & Anatomy, 1027 E 57th Street, Chicago, IL 60637, USA

*corresponding author

Published March 2023 in Molecular Biology and Evolution

Volume 40, Issue 3, March 2023, msad037, https://doi.org/10.1093/molbev/msad037

Abstract

Despite an abundance of gene expression surveys, comparatively little is known about Hox gene function in Chelicerata. Previous investigations of paralogs of *labial* (*lab*) and *Deformed* (*Dfd*) in a spider have shown that these play a role in tissue maintenance of the pedipalpal segment (*lab-1*) and in patterning the first walking leg identity (*Dfd-1*), respectively. However, extrapolations of these data across chelicerates are hindered by the existence of duplicated Hox genes in arachnopulmonates (e.g., spiders and scorpions), which have resulted from an ancient whole genome duplication event. Here, we investigated the function of the single-copy ortholog of *lab* in the harvestman *Phalangium opilio*, an exemplar of a lineage that was not subject to this whole genome duplication. Embryonic RNAi against *lab* resulted in two classes of phenotypes: homeotic transformations of pedipalps to chelicerae, as well as reduction and fusion of the pedipalpal and leg 1 segments. To test for combinatorial function, we performed double knockdown of *lab* and *Dfd*, which resulted in a homeotic transformation of both pedipalps and first walking legs into cheliceral identity, whereas the second walking leg is transformed into a pedipalpal identity. Taken together, these results elucidate a model for the Hox logic of head segments in Chelicerata. To substantiate the validity of this model, we performed expression surveys for *lab* and *Dfd* paralogs in scorpions and horseshoe crabs. We show that repetition of morphologically similar appendages is correlated with uniform expression levels of the Hox genes *lab* and *Dfd*, irrespective of the number of gene copies.

Keywords: Arthropoda | serial homology | Opiliones | tritocerebrum | Xiphosura | Hox1

Introduction

The study of body plan evolution has a rich history, punctuated by advances in comparative embryology, genomics, and functional approaches. Arthropod body plan evolution has served as a particularly fertile proving ground for numerous concepts in evolutionary developmental biology, such as tests of serial homology, cooption of regulatory networks, evolution of cis-regulatory elements, and developmental systems drift (e.g (Auman and Chipman, 2018; Gompel et al., 2005; Jager et al., 2006; Oliver et al., 2014; Setton and Sharma, 2018)). Among these central contributions is the mechanistic understanding of arthropod Hox genes, a family of conserved transcription factors that play a role in regionalizing the antero-posterior axis across Bilateria. Numerous expression surveys have highlighted conserved and labile aspects of Hox expression domains across the arthropod tree of life, and advances in functional toolkits for emerging model species have pinpointed both canonical and non-canonical roles of these genes in body plan patterning (e.g (Abzhanov and Kaufman, 1999; Brena et al., 2006; Hughes and Kaufman, 2002a; Mahfooz et al., 2007; Martin et al., 2016; Medved et al., 2015)). However, the availability of functional datasets for Hox genes is highly asymmetrical, with much of what is known about anterior Hox genes stemming from pancrustacean (and particularly, insect) datasets. Whereas fundamentals of Hox gene dynamics are well understood in pancrustacean models like the fruit fly *Drosophila melanogaster* and the amphipod *Parhyale hawaiensis*, the functions of their homologs remain poorly understood in Chelicerata (e.g., sea spiders, horseshoe crabs, arachnids), the sister group to the remaining arthropods.

The bauplan of Chelicerata—and specifically, of arachnids—generally consists of two tagmata, the anterior prosoma (which bears the mouthparts and walking legs) and the

posterior opisthosoma (which bears modified appendages or may lack appendages altogether in different groups) (Sharma et al., 2014b). The mouthparts of chelicerates typically consist of a pair of chelicerae, innervated by the deutocerebrum (middle region of the tripartite brain; (Damen et al., 1998; Telford and Thomas, 1998); and a pair of pedipalps, innervated by the tritocerebrum (posterior region of the tripartite brain). Posterior to these are segments that bear four pairs of legs, which are variably modified across chelicerate diversity. Modifications of this general architecture are found in Pycnogonida (sea spiders), which lack an opisthosoma and may bear additional leg pairs (e.g., the ovigers; ten- and twelve-legged species in some genera) (Arnaud and Bamber, 1988; Ballesteros et al., 2021). Additionally, Xiphosura (the horseshoe crabs) exhibit anatomically identical pedipalps and walking legs with respect to the number of podomeres (leg segments) (Shultz, 1989; Snodgrass, 1952); only the last pair of walking legs (the pusher leg) exhibits a distinct morphology, owing to its larger size, leaf-like terminal ornamentation, and the presence of the flabellum (an exite; see discussion in (Bruce, 2021)). Whereas the pedipalp (tritocerebral appendage) of Xiphosura is modified in adult males (comparable to the pedipalps of spiders or the third walking legs of Ricinulei), in females, this appendage is indistinguishable from the walking legs in the three immediately posterior segments, barring minor differences in size. As a result, Xiphosura is variably described in the literature as having pedipalps and four pairs of legs, or as having five pairs of legs (i.e segmental homonomy) (Shultz, 1989; Snodgrass, 1952).

Functional datasets informing the patterning of chelicerate prosomal segments have long remained fragmentary. In the spider *Parasteatoda tepidariorum*, it was previously shown that one copy of the Hox gene *labial* (termed *labial-1*) is necessary for the maintenance of the pedipalpal and L1 segment; maternal RNAi against this gene resulted in the reduction or complete deletion of the tritocerebral segment (and occasionally, also the L1 segment;

(Pechmann et al., 2015). The implied function of segment maintenance is contrary to the broadly understood canonical role of Hox genes as drivers of homeosis. By contrast, RNAi against a copy of *Deformed* (termed *Deformed-1*; abbr. *Dfd-1*) in the same species resulted in homeotic transformation of the first walking leg into a pedipalpal identity, as inferred from both the morphology of the transformed appendage and the increased expression of *lab-1* in the ectopic pedipalp (Pechmann et al., 2015). No functional outcomes were reported from RNAi against the paralogs of each gene (i.e., spider *lab-2* and *Dfd-2*). The duplicates of these Hox genes are attributable to a shared genome duplication in the common ancestor of Arachnopulmonata, a group of six orders that includes spiders, scorpions, and pseudoscorpions (Ontano et al., 2021; Schwager et al., 2017; Sharma et al., 2014a). Genomic datasets have shown that nearly all 20 Hox duplicates have been retained across arachnopulmonate lineages, paralleling the evolutionary history of the vertebrates (Gainett and Sharma, 2020; Harper et al., 2021; Leite et al., 2018; Ontano et al., 2021).

However, as these duplicates are restricted to a subset of arachnid orders, it is not clear how well their dynamics reflect the ancestral condition of single-copy patterning genes, particularly given the possibility of subfunctionalization or neofunctionalization of new gene copies in arachnopulmonate exemplars (Benton et al., 2016; Sharma et al., 2014b; Turetzek et al., 2015; Turetzek et al., 2017). In addition, their incidence and retention in models like *P. tepidariorum* raises the specter of functional redundancy and compensatory effects between paralogs in RNAi experiments. For this reason, functional datasets from chelicerate taxa that did not undergo genome duplication events are especially valuable for comparison. Accordingly, recently comparative studies have focused on the unprepossessing harvestman species *Phalangium opilio*, which exhibits an unduplicated genome and is amenable to gene silencing techniques (Gainett et al., 2021; Sharma et al., 2012a; Sharma et al., 2012b; Sharma

et al., 2013). Functional datasets in the harvestman *Phalangium opilio* have shown that the single-copy *Dfd* ortholog is required for patterning the identity of the first two pairs of walking legs; knockdown of *Dfd* resulted in the homeotic transformation of both L1 and L2 legs (first and second walking leg pairs) into pedipalps. Whereas RNAi against *Sex combs reduced* (*Scr*) had no effect by itself, double knockdown of *Dfd* and *Scr* resulted in homeotic transformations of L1-L3 legs toward pedipalpal identity (Gainett et al., 2021). To date, no other functional data are available for the Hox genes of this species.

Here, we focused on discovering the functions of single-copy orthologs of anterior Hox genes in *P. opilio*, with the aim of understanding the Hox logic of the arachnid head segments. We show that knockdown of *lab* results in homeotic transformation of the pedipalp into cheliceral identity, whereas double knockdown of both *lab* and *Dfd* transforms both pedipalps and first walking legs into chelicerae. Having established a ground plan for the patterning of segmental identities in the harvestman, we explored the expression dynamics of *lab* and *Dfd* paralogs in a scorpion and a horseshoe crab, as points of comparison to spider and harvestman Hox genes. This survey substantiates a correlation between heterogeneity of Hox expression levels and the morphological disparity of prosomal appendages.

Results

Expression of harvestman labial

The single copy *Popi-labial* (*Popi-lab*) is expressed as a ring in the equator of the embryo prior to the formation of the germ band, and in dispersed cells towards the center of the encircled labial domain (Fig. 1a–a''). In the stage where the antero-posterior axis forms and the anterior tagma (prosoma) becomes segmented, strong expression localizes to the entire tritocerebral and L1 segment, as marked by the segment polarity gene *Popi-en* (Fig. 1b–b''). *Popi-lab* is also expressed in the posterior part of the L2 and L3 segments (Fig. 1b). During formation of limb buds, *Popi-lab* is strongly expressed in the ectoderm of the pedipalp and L1 limb buds, and in the mesoderm of L2 and L3 nascent limb buds (Fig. 1c–d''). With the sequential addition of posterior body segments, additional paired dots of expression appear in sequence adjacent to the ventral midline (Fig. 1e–g'', arrows). Two antero-posterior stripes of expression occur in the ventral ectoderm of the L2 segment (Fig. 1c–g"). A sharp anterior boundary of expression in the tritocerebral segment, as well as strong expression in the tritocerebral and L1 appendages and in the ventral ectoderm of these segments, are maintained throughout the stages investigated. (Fig. 1c–g'').

Knockdown of lab *results in homeotic pedipalp-to-chelicera transformations and defects in the tritocerebral segment*

To investigate the function of *Popi-lab*, we conducted RNA interference (RNAi) via embryonic injections of double-stranded RNA (dsRNA). *Popi-lab* RNAi hatchlings presented a spectrum of defects affecting the tritocerebral segment. The chelicera and leg identities

were unaffected in hatchlings, whereas the pedipalps were transformed into cheliceral identity (n=35/74) (Fig. 2a-j; Supplementary File, Fig. S1, S2), as evidenced by a proximal podomere and a distal podomere with two claws (chela) (Fig 2b, g, l). Milder transformation consisted of partial truncation of the pedipalpal podomeres and unaffected claw (Fig. 2h, insets), whereas stronger homeosis consisted of a tritocerebral appendage with proximal podomere and an ectopic claw in the distal podomere (Fig. 2h, insets). We note that the most distal part of the transformed appendages did not completely transform into a chela, as the claws retained an identity similar to the pedipalpal and leg claws. Homeosis in the embryonic appendages was evidenced by the branching and reduced length of the tritocerebral limb (Supplementary File, Fig. S1 a–b).

In addition to homeosis, a subset of *Popi-lab* RNAi hatchlings (11%; n=4/35) exhibited defects in the segment boundaries of the tritocerebral segment, as evidenced by the fusion of the tritocerebral appendage to either the chelicera or to L1 (Fig. 3a–f; Supplementary File, Fig. S3). *Popi-lab* RNAi embryos assayed for the segmentation marker *Popi-engrailed* (*Popi-en*) revealed a reduction in the tritocerebral and L1 segment relative to other prosomal segments, and juxtaposition of consecutive *Popi-en* expression stripes (Supplementary File, Fig. S3). Taken together, these results suggest that in *P. opilio labial* is necessary for both of the specification of the tritocerebral segment and for the identity of the pedipalp.

Double knockdown of lab + Dfd *results in transformations of pedipalps and L1 legs into chelicerae*

While RNAi against *lab* results in transformation to cheliceral identity in the pedipalpal segment, we observed that the identity of L1 was not comparably affected, despite the

presence and maintenance of *lab* transcript abundance in L1 territory in the wild type condition. We postulated that that absence of homeotic function in L1 may be attributable to functional redundancy with another anteriorly expressed Hox gene such as *Dfd* and *proboscipedia* (*pb*), though the latter gene is only expressed in subdomains of the prosomal appendages (Sharma et al., 2012a). To test this possibility, we conducted RNAi against *proboscipedia* (*pb*) and double RNAi against *Popi-lab* and *Popi-Dfd*. Preliminary efforts with RNAi against *pb* resulted in no observable phenotype (data not shown); this outcome, together with the restriction of *pb* expression to only some of the tissue from the pedipalpal through L4 segment (Sharma et al., 2012a), suggested that *pb* may not act as a canonical Hox gene in *P. opilio*.

Double RNAi against *Popi-lab* + *Popi-Dfd* resulted in homeotic transformations (n=43/76) affecting tritocerebral, L1 and L2 segments (Fig. 2k–o). Fusions of the tritocerebral appendage with chelicera or L1 leg (n=9/76) were also observed, accompanied by homeosis in the pedipalp, L1 leg, or both appendages (Fig. 3g and h). A subset of the hatchlings (n=22/43) exhibit an additive phenotype, that is, the combination of both pedipalp-to-chelicera (as in single *Popi-lab* RNAi) and leg-to-pedipalp (as in single *Popi-Dfd* RNAi (Gainett et al., 2021)), as evidenced by the truncation of the tritocerebral appendage, and loss of the metatarsus in L1 leg. Notably, the effect in the legs in these cases is similar to the weak phenotypes of single *Popi-Dfd* RNAi (Gainett et al., 2021), with a partial transformation of L1 and L2 legs into pedipalps. A smaller subset of hatchlings exhibited either single *Popi-lab* RNAi phenotype (1/43) or single *Popi-Dfd* RNAi (3/43) phenotype (Supplementary Fig. S2). A new synergistic phenotype was also observed (17/43), in which both pedipalp-to chelicera and the L1 leg-to-chelicera homeosis occurred (Fig. 2k-o; Supplementary Fig. S2). Notably,

we observed only partial L2 leg-to-pedipalp transformations, as evidenced by the lack of metatarsus (Fig. 2o).

These results suggest that both *lab* and *Dfd* repress cheliceral identity, and are consistent with a posterior prevalence model wherein *lab* is required for patterning pedipalpal identity, and *Dfd* for patterning leg identity.

Expression of lab *and* Dfd *duplicates in a scorpion and a horseshoe crab*

A notable complexity in chelicerate Hox cluster evolution is the occurrence of paralogs incurrent from whole genome duplications, which parallels the history of the vertebrates (Dehal and Boore, 2005; Wagner et al., 2003). Furthermore, the whole genome duplications in Arachnopulmonata (spiders, scorpions, and four other arachnid orders) and in Xiphosura (horseshoe crabs) are inferred to have occurred independently (Ontano et al., 2021; Schwager et al., 2017; Shingate et al., 2020a). Hox gene copies are of interest in comparative studies because they may undergo subfunctionalization or neofunctionalization, which in turn may correlate with diversification of body plans (Schwager et al., 2017; Sharma et al., 2014b; Wagner et al., 2003). To test the validity of the Hox logic established herein, as well as obtain further insights into the fate of *lab* paralogs in Chelicerata, we investigated the expression patterns of *lab* paralogs in the scorpion *Centruroides sculpturatus* and in the horseshoe crab *Limulus polyphemus*.

The two *labial* paralogs of *C. sculpturatus* (*Cscu-lab-1*, *Cscu-lab-2*) have a clear anterior boundary of expression in the tritocerebral segment, and strong expression in the tritocerebral appendage (Fig. 4a–a’’). The two paralogs have complex patterns of expression on the ventral midline and developing nervous system, with several overlapping domains (Fig. 4a–

a’’). The two *Dfd* paralogs (*Cscu-Dfd-1*, *Cscu-Dfd-2* have an anterior boundary of expression in the fourth head segment (L1) (Fig. 5 a–a’’). *Cscu-Dfd-1* is uniformly expressed in the L1–L4 legs and the ventral neuromeres (Fig. 5 a–a’’). *Cscu-Dfd-2* has largely overlapping expression with respect to its paralog; it differs from *Cscu-Dfd-1* in that it is more strongly expressed in the tips of the legs, it is slightly more anteriorly expressed in the L1 neuromere, and its expression in the opisthosoma is ubiquitous (Fig. 5 a–a’’). Therefore, the unique identity of the scorpion pedipalp with respect to the legs correlates with the strong expression of both *labial* paralogs on this appendage, whereas the identity of the four pairs of legs correlates with uniform expression of both *Deformed* paralogs in all legs.

Four *labial* paralogs (*Lpol-lab-A*, *Lpol-lab-B*, *Lpol-lab-D*, *Lpol-lab-E*) are present in the genome of *L. polyphemus* (five paralogs occur in *Tachypleus gigas* and *Carcinoscorpius rotundicauda*) (Kenny et al., 2015; Shingate et al., 2020a; Shingate et al., 2020b). *Lpol-lab E* has an anterior boundary of expression in the tritocerebral segment, being expressed in the tritocerebral appendage and posterior prosomal appendages throughout embryonic stages analysed (Fig. 4b–d). *Lpol-lab-D* has a largely overlapping expression domain with respect to *Lpol-lab-E*, with a sharp anterior boundary in the tritocerebral segment (Fig. 4 d’), but slightly more distally expressed in the appendages than *Lpol-lab-E*. The short length of *Lpol-lab-A* CDS (469bp) precluded reliable detection of transcripts via HCR in situ hybridization. We did not analyze the putative homolog *Lpol-lab-B*, given its unusual sequence and ambiguous annotation (Electronic Supplementary Methods).

Five *Dfd* paralogs occur in *L. polyphemus* (*Lpol-Dfd-A–E*). *Lpol-Dfd-B*, *Lpol-Dfd-C*, and *Lpol-Dfd-D* have a clear anterior boundary of expression in the L1 segment (Fig. 5 b–d’) and are expressed in the L1–L4 legs. The three paralogs have largely overlapping expression

domains, with the exception that *Lpol-Dfd-B* expression extends to the lateral margin of the germ band and more posteriorly into the opisthosoma at the germband stage (Fig. 5 b). We were not able to detect expression of *Lpol-Dfd-A* and *Lpol-Dfd-E*. Therefore, the shared morphology of the tritocerebral appendage and the subsequent prosomal appendage pairs (legs) correlates with a uniform expression of *labial* paralogs across all post-cheliceral prosomal appendages, and largely overlapping *lab* and *Dfd* paralog expression across the L1-L4 legs.

Discussion

labial *function in Arthropoda*

In the fruit fly *Drosophila melanogaster*, *lab* is expressed in the tritocerebral segment (the intercalary segment in insects, which is reduced and lacks appendages) during embryogenesis and is necessary for normal head involution (Diederich et al., 1989; Merrill et al., 1989). However, there is no reported homeotic function during embryogenesis, which is also the case for a beetle and a milkweed bug (Angelini et al., 2005; Posnien and Bucher, 2010; Schaeper et al., 2010). Homeosis in *lab* mutant flies is only detected in adults, as a shift from head-to-thorax identity of the bristles (Merrill et al., 1989). This homeosis contrasts with the archetypal transformation towards anterior identity upon disruption of most Hox genes.

By contrast, the tritocerebral segment of non-insect (and non-myriapod) arthropods (i.e., Chelicerata and the crustaceans) bears a pair of appendages that express *lab* (Abzhanov and Kaufman, 1999; Jager et al., 2006; Serano et al., 2016; Sharma et al., 2012a; Telford and Thomas, 1998). The only functional data point available for these groups, in the spider *Parasteatoda tepidariorum*, demonstrated a role in tritocerebral (and in some cases, L1) tissue maintenance, similarly to insects, but no homeotic function in tritocerebral appendage identity upon *lab-1* knockdown (Pechmann et al., 2015). In that work, knockdown of *Dfd-1* was also shown to result in homeotic transformation of L1 to pedipalps, with the ectopic pedipalps strongly expressing *lab-1* (Pechmann et al., 2015). This result is suggestive of the possibility that *lab* may play a role as a pedipalpal determinant, but the evidence is indirect, as homeosis induces transformation of the entire pedipalpal program and its associated gene expression patterns.

On the other hand, the spectrum of embryonic defects in *P. opilio* upon *lab* RNAi suggests a dual role for this gene, both in tritocerebral and L1 segment specification (segmental fusions in the phenotypic spectrum) and in canonical patterning of segmental identity (homeotic transformations of pedipalpal to cheliceral identity). These data reconcile the gap between the unexpected loss of the tritocerebral segment in spiders upon RNAi against *lab-1* and the previously unknown identity of the Hox gene that functions as the pedipalpal selector. It is possible that the *lab-1* segmentation phenotype in the spider reflects the severe end of the phenotypic spectrum, as a consequence of the mode and timing of delivery (maternal RNAi). Such a dosage-dependent mechanism could potentially be tested via embryonic RNAi against *lab-1* in *P. tepidariorum*, with the prediction that lower concentrations of dsRNA injected at a later point in embryogenesis may elicit the homeotic phenotype.

More broadly, the significance of the *P. opilio lab* RNAi phenotype is that it represents the only known case of a homeotic function for *lab* aside from *D. melanogaster*. Given the phylogenetic relationship of these two species (a chelicerate and a hexapod), our results suggest that *lab* may play a role as a conserved tritocerebral selector across Arthropoda.

The relationship between diminution of Hox expression and segmental deletions

Severe loss-of-function phenotypes of *lab* homologs exhibit tritocerebral segment loss and/or fusion in two chelicerates (the spider *P. tepidariorum* (Pechmann et al., 2015) and the harvestman *P. opilio*). Increased cell death in *lab* loss-of-function phenotypes has been reported in the spider *P. tepidariorum* (overall in the prosoma) and in the beetle *T. castaneum* (particularly in the tritocerebral segment in early embryogenesis) (Pechmann et al., 2015;

Schaeper et al., 2010), suggesting that *labial* may directly or indirectly regulate anterior segment specification by controlling apoptosis. Similarly, in *D. melanogaster* the Hox genes *Dfd* and *Abdominal-B* have been shown to promote cell death, which shapes segment boundaries (Lohmann et al., 2002). These observations substantiate the possibility that loss or reduction of a Hox gene may be a potential mechanism for some segmental deletions across Panarthropoda. This inference contrasts with the archetypal role of Hox genes as selector genes that confer segmental identity. Deletion of Hox genes from the Hox cluster has been linked to segmental reduction in multiple panarthropod taxa, such as acariform mites (loss of segmentation posterior to the second opisthosomal segment, in tandem with loss of *abdominal-A* (*abd-A*) (Barnett and Thomas, 2013; Grbić et al., 2011)) and tardigrades (loss of a large intermediate region of the antero-posterior axis, in tandem with multiple Hox genes (Smith et al., 2016)). Similarly, in sea spiders, the absence of a segmented opisthosoma has been tentatively linked to atypical patterns of sequence evolution in *abd-A* (Manuel et al., 2006), but expression patterns of posterior Hox genes are entirely missing in this group (Jager et al., 2006). The mechanism of Hox gene pseudogenization and loss was thought to be released selection in the wake of segmental deletions, resulting in accumulation of mutations and eventual loss of nonfunctional genes associated with the deleted regions—Hox gene loss was understood to reflect the consequence of segmental loss, and not the cause (Barnett and Thomas, 2013; Smith et al., 2016).

Nevertheless, it has been recently reported that diminished *Ubx* and *abd-A* expression in the ant *Camponotus floridanus* results in posterior body truncations, which has been linked to a re-wiring of these genes to a lineage-specific function in early germ cell patterning (Rafiqi et al., 2020). In addition, several members of the homeobox gene family are classically known to be key to segmentation across Arthropoda, such as *even-skipped* and *caudal* (Hughes and

Kaufman, 2002b; Mito et al., 2007; Schönauer et al., 2016). Among these are the *Hox3* paralog *bicoid* in cyclorrhaphan flies and *fushi tarazu* (*ftz*) in a subset of insects. Indeed, in Arthropoda, evidence for a canonical Hox function is entirely absent for *Hox3* (Stauber et al., 1999) and greatly limited for *ftz* (Löhr et al., 2001). Our results support the possibility that diminution or loss of *lab* expression may underlie the independent origins of the appendage-free and highly reduced intercalary segment of hexapods and myriapods, as previously articulated by Pechmann et al. (Pechmann et al., 2015). Given the absence of advanced functional toolkits for either the spider or the harvestman, tests of this hypothesis could capitalize upon the availability of genome editing tools in the amphipod crustacean *Parhyale hawaiensis* (Martin et al., 2016; Serano et al., 2016), with the goal of understanding how fine-tuning of *lab* expression regulates the transition between a homeotic transformation and a segmental deletion. More broadly, gene silencing tools developed in the tardigrade model *Hypsibius exemplaris* could be leveraged to test whether a non-canonical role of *lab* as a segmentation gene is an ancestral feature of Panarthropoda (Smith et al., 2016; Tenlen et al., 2013).

Hox logic and evolution of the chelicerate prosoma

Leg identity in arachnids has been previously shown to require the expression of *Dfd*, as knockdown of *Dfd* results in leg-to-pedipalp homeotic transformation in the daddy-longlegs and in a spider (Gainett et al., 2021; Pechmann et al., 2015), with a redundant role of *Scr* in the case of L3 (Gainett et al., 2021). Similarly, it was previously shown that *homothorax* is required in the absence of Hox input for cheliceral identity (Sharma et al., 2015). The RNAi experiments with *lab* (and both *lab* and *Dfd*) presented herein thus enable the completion of a model for anterior prosomal fate specification in the harvestman (Fig. 6), wherein cheliceral

identity is specified by *homothorax* in the absence of Hox input; pedipalpal identity requires high levels of *lab* and low levels of *Dfd*; and leg identity requires *Dfd* expression, with a redundant role of *Scr* at least in the case of the L3 appendage. The genetic basis for L4 identity remains unknown, but is likely attributable to *ftz*, which is strongly expressed in this territory in the manner of a canonical Hox gene (Sharma et al., 2012a).

This refined understanding of anterior Hox function in this chelicerate led us to revisit patterns of morphological disparity across the prosomal segments of Chelicerata more broadly. In the harvestman, each prosomal segment exhibits a distinguishable identity; in the walking legs, this is observed as unique counts of tarsomeres (tarsal articles) associated with each walking leg pair. By contrast, the walking legs of spiders and scorpions are morphologically more similar, being mainly distinguished by allometric differences (e.g., walking legs become longer towards the posterior in scorpions; the first pair of walking legs is the longest in most spiders). Yet, like the harvestman, the pedipalps of both spiders and scorpions are clearly morphologically distinct. Horseshoe crabs exhibit another condition still, with no morphological distinction between the tritocerebral appendages (or "pedipalps") and the posterior three leg pairs, except in adult males (where the tritocerebral appendage is modified into a clasper for reproduction). Only the posterior-most pusher leg has a distinct, unique morphology by comparison to anterior leg pairs in both sexes of horseshoe crabs (Snodgrass, 1952).

Our gene expression surveys revealed an intriguing correlation between the presence of a morphologically distinct tritocerebral segment identity (e.g., pedipalp) and the concentration of *lab* expression in this segment (Fig. 6a). The only arthropod surveyed to date that has strong, uniform *lab* expression on all post-deutocerebral appendages is the horseshoe crab *L.*

polyphemus. This condition correlates with the morphological similarity of their tritocerebral appendage and post-tritocerebral prosomal appendages (i.e. these five appendage pairs have the same number of segments and the first four are indistinguishable in females) (Fig. 6a). The only comparable case is the protonymphon larvae of sea spiders (Pycnogonida), where *lab* is uniformly expressed in the third and fourth head segments, which bear anatomically identical larval appendages at this developmental stage (Jager et al., 2006).

In contrast to the horseshoe crab and sea spider larvae, the *lab* homologs of all other arthropods with a morphologically distinct tritocerebral identity (second antenna, pedipalps, or intercalary segment) exhibit anteriorly restricted expression domains, with highest concentration of expression levels in the tritocerebral (and in some cases, also the fourth head) segment (Fig. 6b–d). This correlation further holds for the onychophoran *Euperipatoides kanangrensis*, in which *lab* is more strongly expressed in the tritocerebral segment and its appendage (and weakly in all posterior segments), correlating with the unique identity of the slime papillae in contrast to posterior lobopods (Fig. 6e) (Eriksson et al., 2010; Janssen et al., 2014).

Furthermore, we discovered that *Dfd* homologs exhibit a similar trend: morphologically similar legs exhibit uniform expression of *Deformed*. Although *Dfd* homologs always bore an anterior boundary in the fourth head segment, expression patterns and transcript levels of *Dfd* copies are highly similar in the L1–L4 segments of spiders (Schwager et al., 2017) and scorpions (this study); and in L1–L4 of horseshoe crabs. This correlation holds upon inclusion of Hox surveys (albeit fragmentary datasets) from other chelicerate groups; uniform expression levels of *Dfd* were detected in the L1–L3 legs of the hexapod larva of an acariform mite (Telford and Thomas, 1998). In contrast to these groups, expression levels of

the single-copy ortholog of *Dfd* in the harvestman are markedly heterogeneous across L1–L4, with a unique expression pattern and level of intensity in each leg (Gainett et al., 2021). The general correlation we observe across these groups is that homonomous blocks of segments (i.e., segments bearing morphologically similar appendages) tend to exhibit similar combinations of Hox expression levels.

These patterns suggest that modifications of pedipalp and leg morphology across Chelicerata may be attributable to modulation of Hox levels, driven by variation in cis-regulation of *lab* and *Dfd* to achieve unique combinations of transcript levels in morphologically distinct appendages. The presence of Hox duplications in Arachnopulmonata, while intriguing from a macroevolutionary perspective, may not be the mechanism that underlies morphological disparity of chelicerate prosomal architecture. We found no evidence that prosomal Hox genes have achieved additional segmental identities in arachnopulmonates by spatial subdivision of paralogs' expression domains (in contrast to opisthosomal Hox copies in spiders and scorpions; (Sharma et al., 2012a; Sharma et al., 2014b). This inference is further underscored by the marked morphological disparity between walking legs across groups of acariform mites, which are comparably diverse to spiders, but are not part of the shared genome duplication at the root of Arachnopulmonata (Ontano et al., 2021; Schwager et al., 2017).

These hypotheses could be tested further by investigating arachnopulmonates with extreme modifications of specific appendage pairs, such as Amblypygi (whip spiders), Uropygi (vinegaroons), and Solifugae (camel spiders), groups that exhibit marked elongation of the first walking leg pair. Recently developed tools for the whip spider *Phrynus marginemaculatus* may prove useful in this regard (Gainett and Sharma, 2020). With respect

to establishment of specific leg-bearing segment identities in the harvestman, advanced tools for misexpression of Hox genes are required in this system for nuanced understanding of Hox transcript combinatorics. Ectopic expression of harvestman *lab* in posterior prosomal segments (L2–L4), for example, is central to testing the hypothesis that co-expression of *lab* and *Dfd* underlies L1 fate specification. We note that such an experiment, if successful in generating four pairs of morphologically identical appendages in *P. opilio*, would strongly accord with the interpretation that homonomous segmentation can evolve secondarily in lineages like Xiphosura—a scenario that is suggested by the phylogenetic distribution of Hox domains across Panarthropoda (Fig. 6), as well as recent phylogenetic work that has recovered horseshoe crabs as derived arachnids (Ballesteros and Sharma, 2019; Ballesteros et al., 2022; Ban et al., 2022)

Materials and Methods

Embryo collection and gene identification

Adult *P. opilio* individuals were collected from Bascom Hill, Madison, WI, USA along the exterior walls of nearby buildings. Gravid females of the scorpion *Centruroides sculpturatus* were hand-collected from sites in Arizona by citizen-scientist collaborators. Embryos of the horseshoe crab *Limulus polyphemus* were collected in Woods Hole, MA, USA in June 2022. Embryonic staging nomenclature for the harvestman and horseshoe crab follows Gainett et al. (Gainett et al., 2022) and Sekiguchi et al. (Sekiguchi et al., 1982), respectivelly. Details of collection, maintenance and fixation of embryos are described in the Electronic Supplementary Methods.

The complete sequences of *P. opilio* Hox genes *labial* (*lab*), *probocipedia* (*pb*), and *Deformed* (*Dfd*) were previously isolated from the *P. opilio* genome (GCA_019434445.1;(Gainett et al., 2021)) (Table S1). The *engrailed* (*en*) ortholog in *P. opilio* was identified by Sharma et al. (Sharma et al., 2012a) (Table S1). The scorpion *C. sculpturatus*, *lab* paralogs (*Cscu-lab-1, Cscu-lab-2)* and Dfd paralogs (*Cscu-Dfd-A, Cscu-Dfd-B)* were identified from the reference genome (GCF_000671375.1; (Schwager et al., 2017)), the A/B nomenclature (here paralogs1 and 2, respectively) of their original description (Schwager et al., 2017). The numerical nomenclature for the scorpion genes was used to differentiate paralogs of the arachnopulmonate whole genome duplication (WGD) from the independently derived Hox paralogs of the horseshoe crab WGDs. Horseshoe crab Hox paralogs are referred here using alphabetic nomenclature, following the convention established in the *Carcinoscorpius rotundicauda* and *Tachypleus gigas* genome assemblies

(Shingate et al., 2020a; Shingate et al., 2020b). The horseshoe crab *L. polyphemus lab* paralogs (*Lpol-lab-A* , *Lpol-lab-B* , *Lpol-lab-D, Lpol-lab-E*) and *Dfd* paralogs (*Lpol-Dfd-A*, *Lpol-Dfd-B*, *Lpol-Dfd-C*, *Lpol-Dfd-D*, *Lpol-Dfd-E*) were retrieved from the reference genome annotation (GCF_000517525.1; (Battelle et al., 2016; Kenny et al., 2015). Accession numbers (Table S2) and phylogenetic analysis of *lab* (Supplementary Fig. S4) and *Dfd* (Supplementary Fig. S5) homologs in the horseshoe crab species are described in the Electronic Supplementary Material.

RNA interference (RNAi) via double-stranded RNA (dsRNA) embryonic injections

Gene cloning and dsRNA synthesis are detailed in the Electronic Supplementary Methods (Table S3). For the single knockdown of *Popi-lab*, two clutches of embryos were injected (n=194). Three clutches were injected in the double knockdown of *Popi-lab* + *Popi-Dfd* (n=217). An additional three clutches were injected with *Popi-lab* dsRNA, and an additional two were injected with *Popi-lab* + *Popi-Dfd* dsRNA which were then fixed and exclusively used in both colorimetric and fluorescent *in situ* hybridization. Two clutches (n=157) were injected with deionized water as a negative control. Injection mixes were prepared with Rhodamine dextran (1:20) for visualization.

Colorimetric and fluorescent in situ *hybridization*

Colorimetric *in situ* hybridization gene expression assays of *P. opilio* used sense (control) and antisense RNA probes labeled with DIG RNA labeling mix (Roche, Basel, Switzerland) and followed published protocols (Sharma et al., 2012a). Images were obtained on a Nikon SMZ25 fluorescent stereomicroscope with a DS-Fi2 digital color camera and driven by

Nikon Elements. Fluorescent *in situ* hybridization followed a modified version of the Molecular Instruments (Los Angeles, CA, USA) hybridization chain reaction (HCR) v.3 protocol (Bruce et al., 2021; Choi et al., 2018). HCR probes for Hox paralogs were designed avoiding regions of high similarity to circumvent non-specific binding. Probe sequences were designed by Molecular Instruments or in an open-source probe design program (Kuehn et al., 2022). Details of probe design and probe sequences are available in the Electronic Supplementary Methods (Table S4–12). Imaging was performed on a Zeiss 710 and Zeiss 780 confocal microscope at the Newcomb Imaging Center, UW-Madison, USA. Z-stacks were projected with maximal intensity mode, and linearly adjusted for brightness and contrast in FIJI (v. 2.9.0/1.53t). Figures were assembled in Adobe Illustrator 2022.

Acknowledgements

We are indebted to Sarah Swanson and the Newcomb Imaging Center (Department of Botany) at the University of Wisconsin-Madison for the imaging infrastructure. Comments from Emily V. W. Setton improved the first draft of the manuscript. We thank two anonymous reviewers and the associate editor for comments on the first draft. This work was supported by the National Science Foundation NSF IOS-2016141 grant to PPS.

Data Availability Statement

The data underlying this article are available in the article and in its online supplementary material.

References

Abzhanov, A. and Kaufman, T. C. (1999). Homeotic genes and the arthropod head: Expression patterns of the *labial*, *proboscipedia*, and *Deformed* genes in crustaceans and insects. *Proc. Natl. Acad. Sci. USA* 96, 10224–10229.

Angelini, D. R., Liu, P. Z., Hughes, C. L. and Kaufman, T. C. (2005). Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). *Dev. Biol.* 287, 440–455.

Arnaud, F. and Bamber, R. N. (1988). The Biology of Pycnogonida. In *Advances in Marine Biology* (ed. Blaxter, J. H. S. and Southward, A. J.), pp. 1–96. Academic Press.

Auman, T. and Chipman, A. D. (2018). Growth zone segmentation in the milkweed bug *Oncopeltus fasciatus* sheds light on the evolution of insect segmentation. *BMC Evol. Biol.* 18, 178.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440–22.

Ballesteros, J. A., Setton, E. V. W., López, C. E. S., Arango, C. P., Brenneis, G., Brix, S., Corbett, K. F., Sánchez, E. C., Dandouch, M., Dilly, G. F., et al. (2021). Phylogenomic resolution of sea spider diversification through integration of multiple data classes. *Mol. Biol. Evol.* 38, 686–701.

Ballesteros, J. A., Santibáñez-López, C. E., Baker, C. M., Benavides, L. R., Cunha, T. J., Gainett, G., Ontano, A. Z., Setton, E. V. W., Arango, C. P., Gavish-Regev, E., et al. (2022). comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Mol. Biol. Evol.* 39, msac021.

Ban, X., Shao, Z., Wu, L., Sun, J. and Xue, X. (2022). Highly diversified mitochondrial genomes provide new evidence for interordinal relationships in the Arachnida. *Cladistics* 38, 452–464.

Barnett, A. A. and Thomas, R. H. (2013). Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *Evodevo* 4, 23–23.

Battelle, B.-A., Ryan, J. F., Kempler, K. E., Saraf, S. R., Marten, C. E., Warren, W. C., Minx, P. J., Montague, M. J., Green, P. J., Schmidt, S. A., et al. (2016). Opsin repertoire and expression patterns in horseshoe crabs: Evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biol. Evol.* 8, 1571–1589.

Benton, M. A., Pechmann, M., Frey, N., Stappert, D., Conrads, K. H., Chen, Y.-T., Stamataki, E., Pavlopoulos, A. and Roth, S. (2016). Toll Genes Have an Ancestral Role in Axis Elongation. *Curr. Biol.* 26, 1609–1615.

Brena, C., Chipman, A. D., Minelli, A. and Akam, M. (2006). Expression of trunk Hox genes in the centipede *Strigamia maritima*: sense and anti-sense transcripts. *Evol. Dev.* 8, 252–265.

Bruce, H. S. (2021). How to align arthropod legs. *Biorxiv* 2021.01.20.427514.

Bruce, H. S., Jerz, G., Kelly, S. R., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D. A., Wolff, C. and Patel, N. H. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol V.1.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A. and Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753.

Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. and Tautz, D. (1998). A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl. Acad. Sci. USA* 95, 10665–10670.

Dehal, P. and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *Plos Biol.* 3, e314.

Diederich, R. J., Merrill, V. K., Pultz, M. A. and Kaufman, T. C. (1989). Isolation, structure, and expression of *labial*, a homeotic gene of the Antennapedia Complex involved in *Drosophila* head development. *Gene Dev.* 3, 399–414.

Eriksson, B. J., Tait, N. N., Budd, G. E., Janssen, R. and Akam, M. (2010). Head patterning and Hox gene expression in an onychophoran and its implications for the arthropod head problem. *Dev. Genes Evol.* 220, 117–122.

Gainett, G. and Sharma, P. P. (2020). Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *EvoDevo* 11, 18–18.

Gainett, G., González, V. L., Ballesteros, J. A., Setton, E. V. W., Baker, C. M., Gargiulo, L. B., Santibáñez-López, C. E., Coddington, J. A. and Sharma, P. P. (2021). The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proc. R. Soc. B* 288, 20211168.

Gainett, G., Crawford, A. R., Klementz, B. C., So, C., Baker, C. M., Setton, E. V. W. and Sharma, P. P. (2022). Eggs to long-legs: embryonic staging of the harvestman *Phalangium opilio* (Opiliones), an emerging model arachnid. *Front. Zool.* 19, 11.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487.

Grbić, M., Leeuwen, T. V., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.

Harper, A., Gonzalez, L. B., Schönauer, A., Janssen, R., Seiter, M., Holzem, M., Arif, S., McGregor, A. P. and Sumner-Rooney, L. (2021). Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates. *G3 Genes Genomes Genet.* 11, jkab299.

Hughes, C. L. and Kaufman, T. C. (2002a). Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* 4, 459–499.

Hughes, C. L. and Kaufman, T. C. (2002b). Exploring myriapod segmentation: the expression patterns of *even-skipped*, *engrailed*, and *wingless* in a centipede. *Dev. Biol.* 247, 47–61.

Jager, M., Murienne, J., Clabaut, C., Deutsch, J., Guyader, H. L. and Manuel, M. (2006). Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider. *Nature* 441, 506–508.

Janssen, R., Eriksson, B. J., Tait, N. N. and Budd, G. E. (2014). Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front. Zool.* 11, 22.

Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., et al. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116, 190–199.

Kuehn, E., Clausen, D. S., Null, R. W., Metzger, B. M., Willis, A. D. and Özpolat, B. D. (2022). Segment number threshold determines juvenile onset of germline cluster expansion in *Platynereis dumerilii*. *J Exp Zoology Part B Mol Dev Evol* 338, 225–240.

Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N.-M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018).

Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* 35, 2240–2253.

Lohmann, I., McGinnis, N., Bodmer, M. and McGinnis, W. (2002). The *Drosophila* Hox gene *Deformed* sculpts head morphology via direct regulation of the apoptosis activator *reaper*. *Cell* 110, 457–466.

Löhr, U., Yussa, M. and Pick, L. (2001). *Drosophila fushi tarazu* a gene on the border of homeotic function. *Curr. Biol.* 11, 1403–1412.

Mahfooz, N., Turchyn, N., Mihajlovic, M., Hrycaj, S. and Popadić, A. (2007). *Ubx* regulates differential enlargement and diversification of insect hind legs. *PLoS One* 2, e866-9.

Manuel, M., Jager, M., Murienne, J., Clabaut, C. and Guyader, H. L. (2006). Hox genes in sea spiders (Pycnogonida) and the homology of arthropod head segments. *Dev. Genes Evol.* 216, 481–491.

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C. and Patel, N. H. (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Curr. Biol.* 26, 14–26.

Medved, V., Marden, J. H., Fescemyer, H. W., Der, J. P., Liu, J., Mahfooz, N. and Popadić, A. (2015). Origin and diversification of wings: Insights from a neopteran insect. *Proc. Natl. Acad. Sci. USA* 112, 15946–15951.

Merrill, V. K. L., Diederich, R. J., Turner, F. R. and Kaufman, T. C. (1989). A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev. Biol.* 135, 376–391.

Mito, T., Kobayashi, C., Sarashina, I., Zhang, H., Shinahara, W., Miyawaki, K., Shinmyo, Y., Ohuchi, H. and Noji, S. (2007). *even-skipped* has gap-like, pair-rule-like, and segmental functions in the cricket *Gryllus bimaculatus*, a basal, intermediate germ insect (Orthoptera). *Dev. Biol.* 303, 202–213.

Oliver, J. C., Beaulieu, J. M., Gall, L. F., Piel, W. H. and Monteiro, A. (2014). Nymphalid eyespot serial homologues originate as a few individualized modules. *Proc. R. Soc. B* 281, 20133262.

Ontano, A. Z., Gainett, G., Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., et al. (2021). Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol. Biol. Evol.* 38, 2446–2467.

Pechmann, M., Schwager, E. E., Turetzek, N. and Prpic, N.-M. (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc. R. Soc. B* 282, 20151162.

Posnien, N. and Bucher, G. (2010). Formation of the insect head involves lateral contribution of the intercalary segment, which depends on *Tc-labial* function. *Dev. Biol.* 338, 107–116.

Rafiqi, A. M., Rajakumar, A. and Abouheif, E. (2020). Origin and elaboration of a major evolutionary transition in individuality. *Nature* 1–30.

Schaeper, N. D., Pechmann, M., Damen, W. G. M., Prpic, N.-M. and Wimmer, E. A. (2010). Evolutionary plasticity of *collier* function in head development of diverse arthropods. *Dev. Biol.* 344, 363–376.

Schönauer, A., Paese, C. L. B., Hilbrant, M., Leite, D. J., Schwager, E. E., Feitosa, N. M., Eibner, C., Damen, W. G. M. and McGregor, A. P. (2016). The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* 143, 2455–2463.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.

Sekiguchi, K., Yamamichi, Y. and Costlow, J. D. (1982). Horseshoe crab developmental studies I. Normal embryonic development of *Limulus polyphemus* compared with *Tachypleus tridentatus*. *Prog. Clin. Biol. Res.* 81, 53–73.

Serano, J. M., Martin, A., Liubicich, D. M., Jarvis, E., Bruce, H. S., La, K., Browne, W. E., Grimwood, J. and Patel, N. H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean *Parhyale hawaiensis*. *Dev. Biol.* 409, 297–309.

Setton, E. V. W. and Sharma, P. P. (2018). Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proc. Natl. Acad. Sci. USA* 128, 201720193–10.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012a). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012b). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* 14, 522–533.

Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol. Dev.* 15, 228–242.

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C. and Giribet, G. (2014a). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31, 2963–2984.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Wheeler, W. C. (2014b). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661–20140661.

Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C. and Extavour, C. G. (2015). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. B* 282, 20150698.

Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E. and Venkatesh, B. (2020a). Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.* 11, 2322.

Shingate, P., Ravi, V., Prasad, A., Tay, B.-H. and Venkatesh, B. (2020b). Chromosome-level genome assembly of the coastal horseshoe crab (*Tachypleus gigas*). *Mol. Ecol. Resour.* 20, 1748–1760.

Shultz, J. W. (1989). Morphology of locomotor appendages in Arachnida: evolutionary trends and phylogenetic implications. *Zool. J. Linn. Soc.* 97, 1–55.

Smith, F. W., Boothby, T. C., Giovannini, I., Rebecchi, L., Jockusch, E. L. and Goldstein, B. (2016). The compact body plan of tardigrades evolved by the loss of a large body region. *Curr. Biol.* 26, 224–229.

Snodgrass, R. E. (1952). Limulus. In *Textbook of Arthropod Anatomy*, pp. 20–40. Cornell University Press.

Stauber, M., Jäckle, H. and Schmidt-Ott, U. (1999). The anterior determinant *bicoid* of *Drosophila* is a derived *Hox* class 3 gene. *Proc. Natl. Acad. Sci. USA* 96, 3786–3789.

Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl. Acad. Sci. USA* 95, 10671–10675.

Tenlen, J. R., McCaskill, S. and Goldstein, B. (2013). RNA interference can be used to disrupt gene function in tardigrades. *Dev. Genes. Evol.* 223, 171–181.

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J. and Prpic, N.-M. (2015). Neofunctionalization of a duplicate *dachshund* gene underlies the evolution of a novel leg segment in arachnids. *Mol. Biol. Evol.* 33, 109–121.

Turetzek, N., Khadjeh, S., Schomburg, C. and Prpic, N.-M. (2017). Rapid diversification of *homothorax* expression patterns after gene duplication in spiders. *BMC Evol. Biol.* 17, 168.

Wagner, G. P., Amemiya, C. and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. USA* 100, 14603–14606.

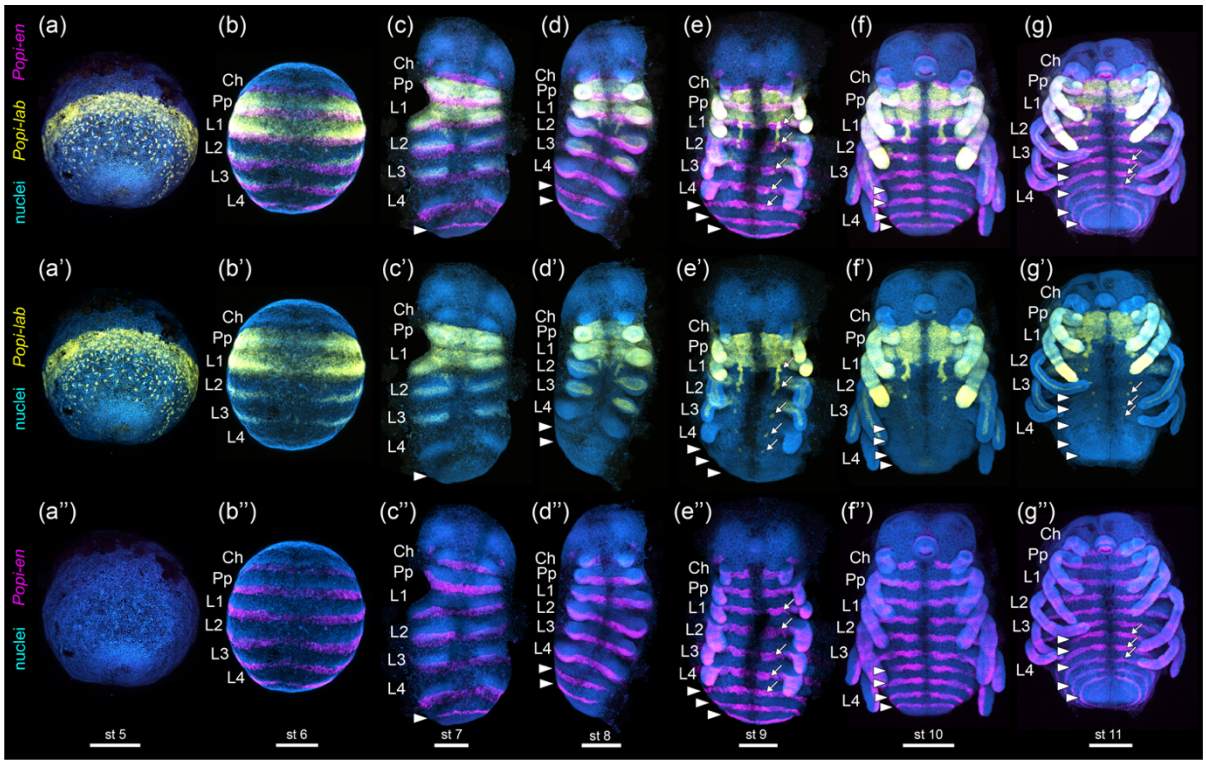


Figure 1: Wild type expression of *Popi-lab* and *Popi-en* across embryonic stages. Anterior at top. (a) Wholemount, ventrolateral view of the germ disc. (b) Wholemount, ventral view. (c–g): Flatmount, ventral view. Nuclei in blue (Hoechst). *Popi-en*: magenta. *Popi-lab*: yellow. (a–g): Hoechst, *Popi-lab* and *Popi-en* multiplexed. (a'–g'): Hoechst and *Popi-lab*, same embryos as previous row. (a''–g''): Hoechst and *Popi-en*, same embryos as previous row. Arrowheads: *Popi-en* stripes marking the posterior of each segment. Arrows: dots of *Popi-lab* expression adjacent to the ventral midline. Ch: chelicera; Pp: pedipalp; L1–L4: L1–L4 legs; st, stage number. Scale bars: 200 μm.

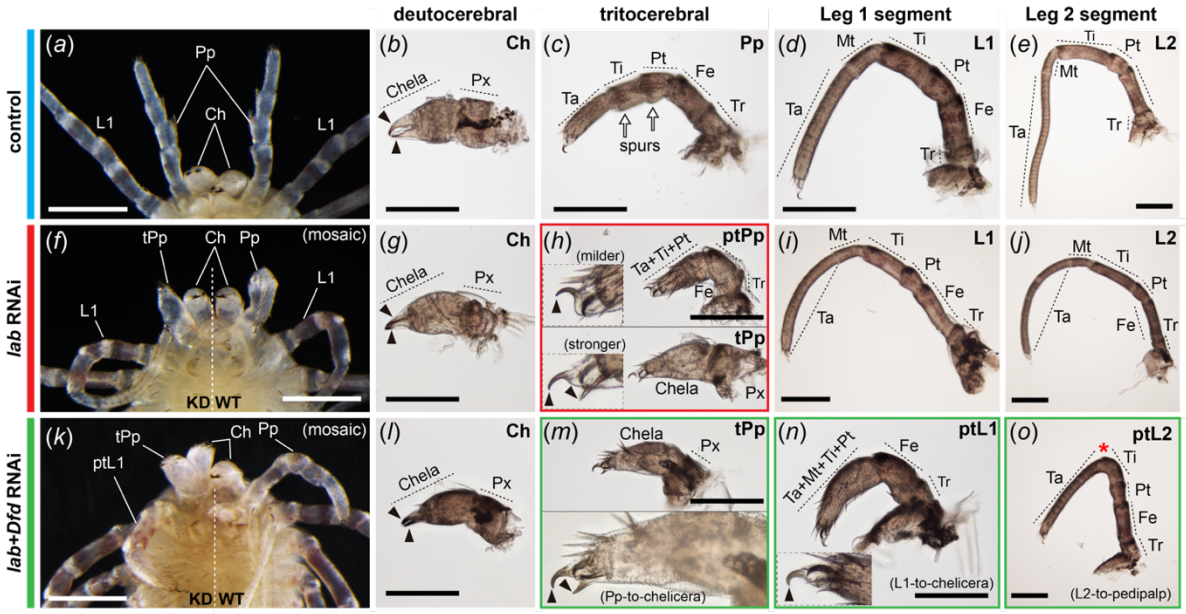


Figure 2: *Popi-lab* knockdown results in homeotic transformations. Brightfield images of hatchlings (postembryos) of control (dH_2O-injected) (a–e), *Popi-lab* RNAi (f–j), and *Popi-lab* + *Popi-Dfd* RNAi (k–o) experiments. Panels (a), (f) and (k) are ventral views of the anterior prosoma. Other panels are dissected appendages in lateral view with proximal at right. (b, g and l): Deutocerebral appendage (chelicera). (c, h, and m): Tritocerebral appendage. (d, i, and n): Appendage of the L1 segment. (e, j, and o): Appendage of the L2 segment. Note that, because the 1st instar's cuticle is already secreted and visible inside the postembryo cuticle, two sets of claws are discernible in some panels. Black arrowhead: claw; red asterisk: missing metatarsus (Mt); white arrows: pedipalpal spurs; Ch: chelicera; Pp: pedipalp; tPp: transformed pedipalp; ptPp: partially transformed pedipalp; L1–L2: L1–L2 appendages; ptL1–ptL2: partially transformed L1–L2 appendages; Ta: tarsus; Mt: metatarsus; Ti: tibia; Pt: patella; Fe: femur; Tr: trochanter; Px: proximal cheliceral segment; WT, wild type side of individual; KD, knockdown side of individual. Scale bars: 200 μm.

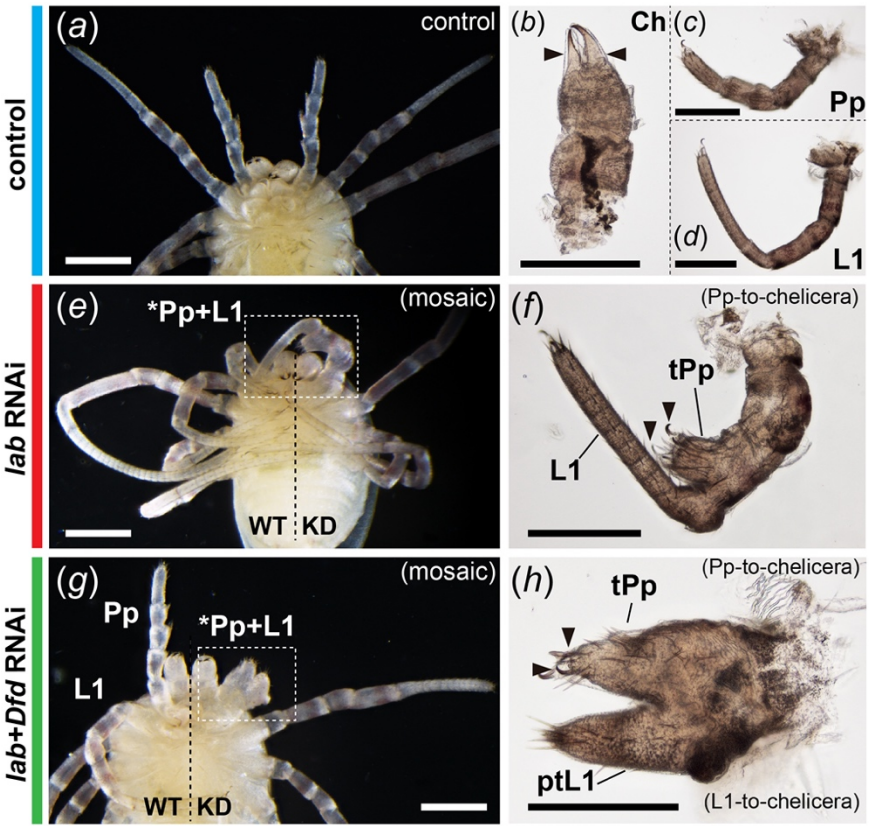


Figure 3: *Popi-lab* knockdown results in fusion of the pedipalp and L1 leg. Brightfield images of hatchlings (postembryos) of control (dH_2O-injected) (a–d), *Popi-lab* RNAi (e–f), and *Popi-lab* + *Popi-Dfd* RNAi (g–h) experiments. Phenotypes in g–h present both homeosis and fusion. Panels (a, e, and g) are ventral views of the anterior prosoma. Other panels are dissected appendages in lateral view. Mosaic individuals in (e) and (g) are affected on the right side of the picture. Black arrowhead: claw; Ch: chelicera; Pp: pedipalp; tPp: transformed pedipalp; L1: L1 appendage; ptL1, partially transformed L1 leg; WT, wildtype side of individual; KD, knockdown side of individual. Scale bars: 200 μm.

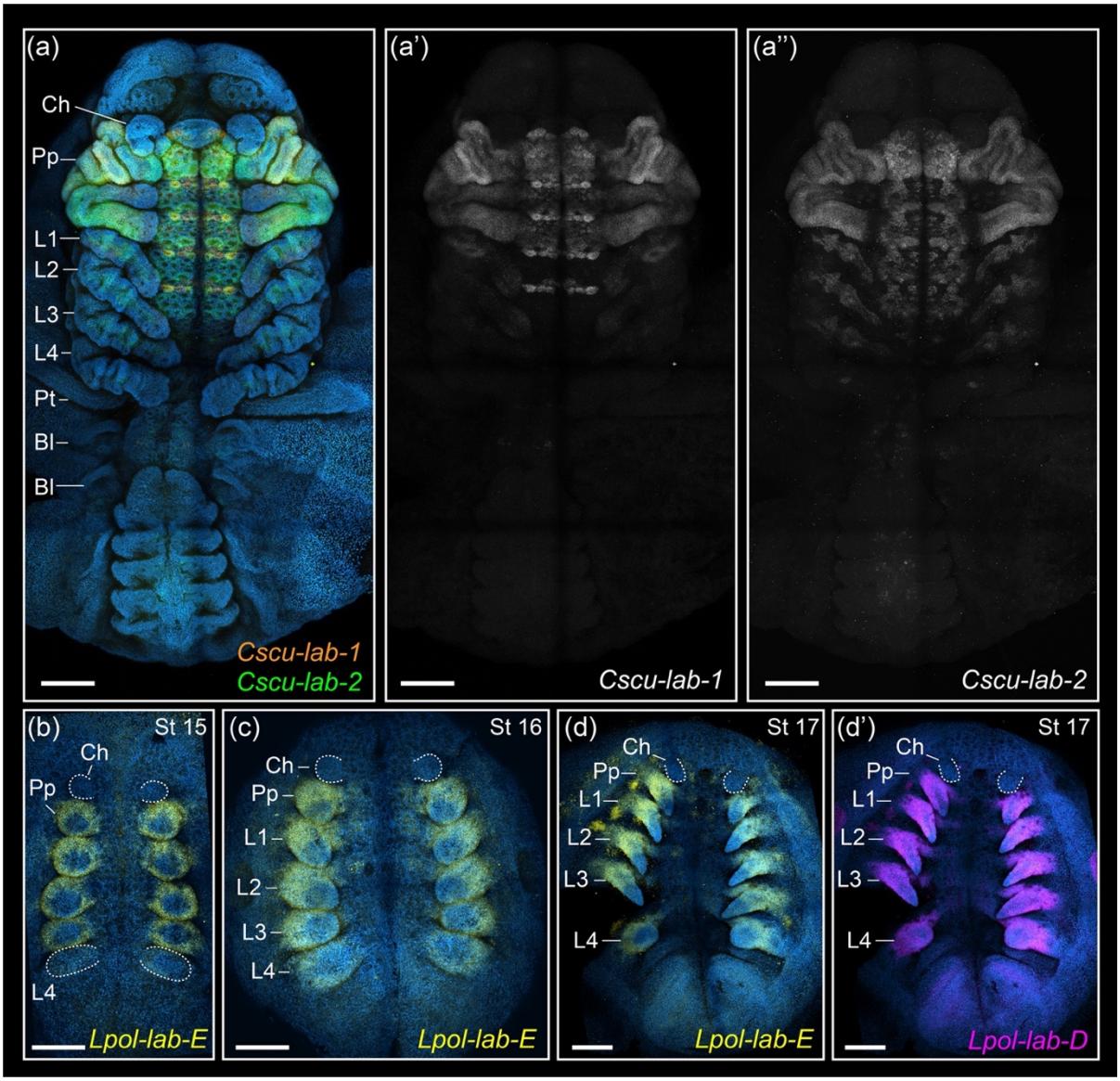


Figure 4. HCR in situ hybridization of *labial* paralogs in the scorpion *Centruroides sculpturatus* (a–a'') and the horseshoe crab *Limulus polyphemus* (b–d'). Maximum intensity projections of flat mounted embryos in ventral view. Hoechst counter staining in blue. Same letters indicate same embryo (multiplexed). (a) Merged projections of *Cscu-lab-1* (orange), *Cscu-lab-2* (green). Overlap of *lab* paralogs appears in yellow. Note that apparent *labial* expression in the posterior chelicera and labrum is in the tritocerebral tissue; overlapping of these regions is incurring by the maximum intensity projection. (a') Single channel projection of *Cscu-lab-1*. (a'') Single channel projection of *Cscu-lab-2*. (b–d) *Lpol-lab-E* expression (yellow). (d'') *Lpol-lab-D* expression (magenta). Ch: chelicera; Pp: pedipalp; L1–L4: L1–L4 legs. Scale bars: 200 μm.

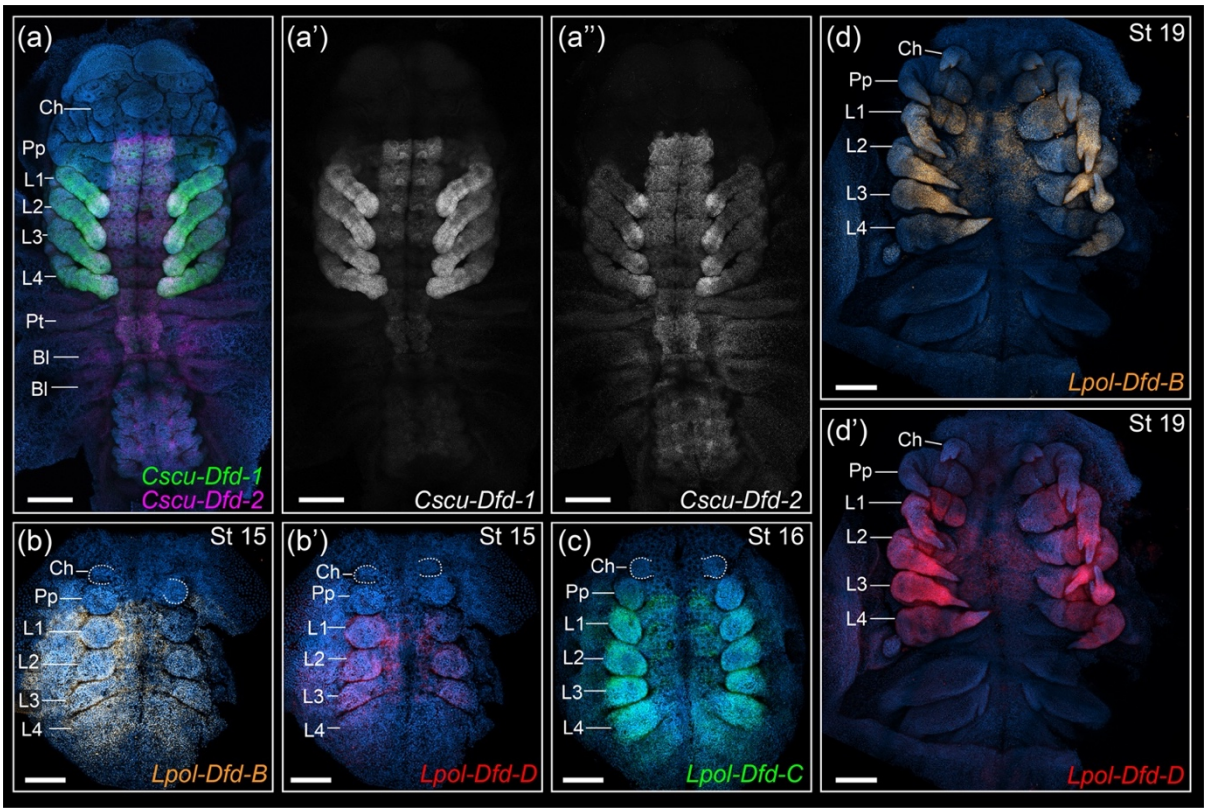


Figure 5: HCR in situ hybridization of *Deformed* paralogs in the scorpion *Centruroides sculpturatus* (a–a'') and the horseshoe crab *Limulus polyphemus* (b–d'). Maximum intensity projections of flat mounted embryos in ventral view. Hoechst counter staining in blue. Same letters indicate same embryo (multiplexed). (a) Merged projections of *Cscu-Dfd-1* (green), *Cscu-Dfd-2* (magenta). (b) *Lpol-Dfd-B* expression (orange). (b') *Lpol-Dfd-D* expression (red). (c) *Lpol-Dfd*-C expression (green). (d) *Lpol-Dfd-B* expression (orange). (d') *Lpol-Dfd-D* expression (red). Ch: chelicera; Pp: pedipalp; L1–L4: L1–L4 legs. Scale bars: 200 μm.

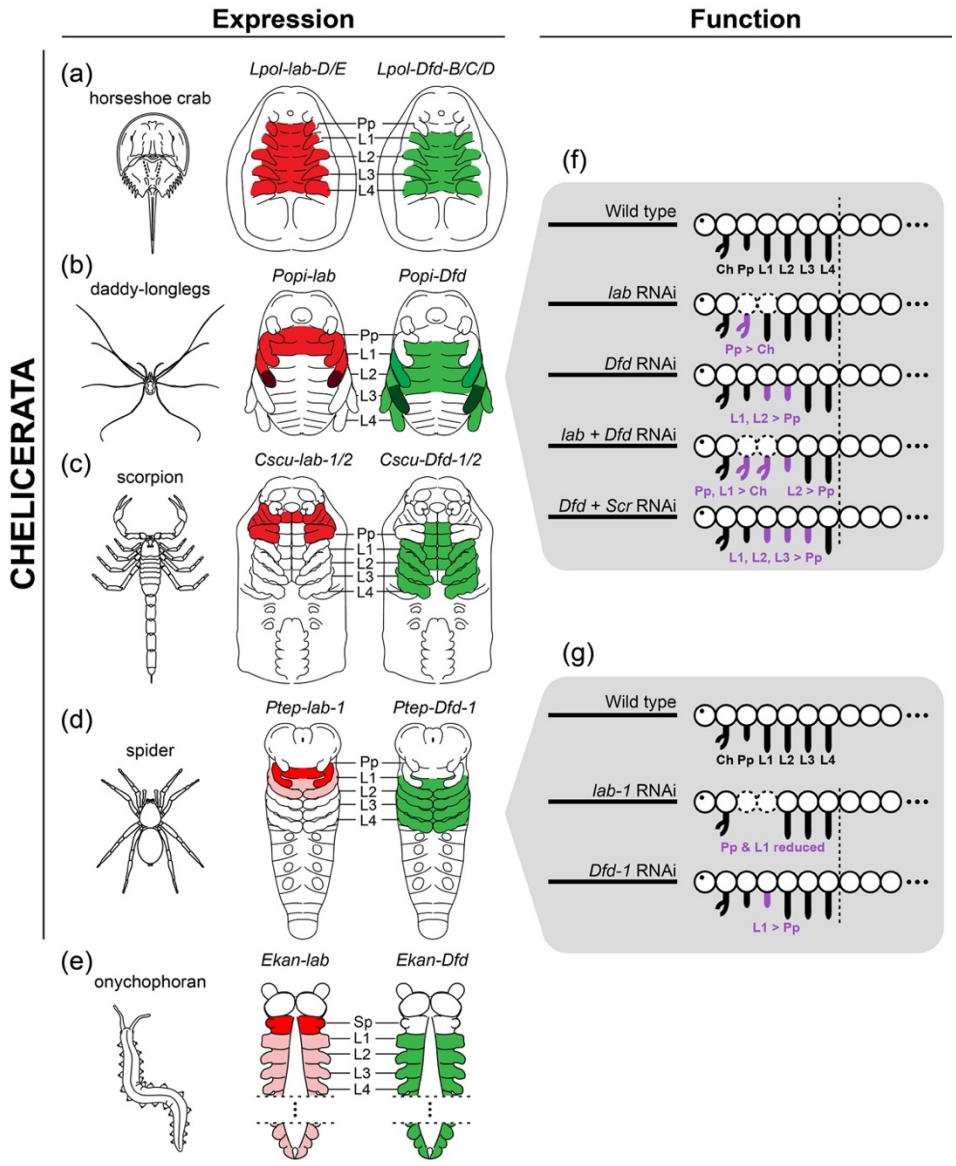


Figure 6: Summary of *labial* (red; left schematics) and *Deformed* (green; right schematics) expression and functional data in Chelicerata and Onychophora (outgroup). Darker shades reflect concentration of expression. (a) Horseshoe crab *Limulus polyphemus*. (b) Daddy-longlegs *Phalangium opilio*, after Sharma et al. (Sharma et al., 2012a), and Gainett et al. (Gainett et al., 2021). (c) Scorpion *Centruroides sculpturatus*. (d) Spider *Parasteatoda tepidariorum*, after Schwager et al. ((Schwager et al., 2017)). (e) Onychophoran *Euperipatoides kanangrensis*, after Eriksson et al (Eriksson et al., 2010) and Janssen et al. (Janssen et al., 2014). (f) Summary of phenotypic outcomes of Hox RNAi experiments in *P. opilio*. Dotted body segments indicate segment reduction. Data for the single knockdown of *Popi-Dfd* and *Popi-Scr* knockdown after Gainett et al. (Gainett et al., 2021). (g) Summary of phenotypic outcomes of *Ptep-lab-1* and *Ptep-Dfd-1* RNAi experiments in *P. tepidariorum*, after Pechmann et al. (Pechmann et al., 2015). Ch, chelicera; Pp, pedipalp; L1–L4, L1–L4 legs; Sp, slime
papilla.

Chapter 4

Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning

Guilherme Gainett and Prashant P. Sharma*

*Correspondence: guilherme.gainett@wisc.edu
Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA

Published August 2020 in EvoDevo

Volume 11, Issue 18, doi.org/10.1186/s13227-020-00163-w

Abstract

Background: The resurgence of interest in the comparative developmental study of chelicerates has led to important insights, such as the discovery of a genome duplication shared by spiders and scorpions, inferred to have occurred in the most recent common ancestor of Arachnopulmonata (a clade comprising the five arachnid orders that bear book lungs). Nonetheless, several arachnid groups remain understudied in the context of development and genomics, such as the order Amblypygi (whip spiders). The phylogenetic position of Amblypygi in Arachnopulmonata posits them as an interesting group to test the incidence of the proposed genome duplication in the common ancestor of Arachnopulmonata, as well as the degree of retention of duplicates over 450 Myr. Moreover, whip spiders have their first pair of walking legs elongated and modified into sensory appendages (a convergence with the antennae of mandibulates), but the genetic patterning of these antenniform legs has never been investigated.

Results: We established genomic resources and protocols for cultivation of embryos and gene expression assays by in situ hybridization to study the development of the whip spider *Phrynus marginemaculatus*. Using embryonic transcriptomes from three species of Amblypygi, we show that the ancestral whip spider exhibited duplications of all ten Hox genes. We deploy these resources to show that paralogs of the leg gap genes *dachshund* and *homothorax* retain arachnopulmonate-specific expression patterns in *P. marginemaculatus*. We characterize the expression of leg gap genes *Distal-less, dachshund-1/2* and *homothorax-1/2* in the embryonic antenniform leg and other appendages, and provide evidence that allometry, and by extension the antenniform leg fate, is specified early in embryogenesis.

Conclusion: This study is the first step in establishing *P. marginemaculatus* as a chelicerate model for modern evolutionary developmental study and provides the first resources sampling whip spiders for comparative genomics. Our results suggest that Amblypygi share a genome duplication with spiders and scorpions and set up a framework to study the genetic specification of antenniform legs. Future efforts to study whip spider development must emphasize the development of tools for functional experiments in *P. marginemaculatus*.

Keywords: Arachnida, Arachnopulmonata, Gene duplication, Paralogs, Leg gap genes, Sensory biology

Background

In the past 25 years, comparative developmental study of chelicerates has resurged with the integration of molecular biology and genomics, spearheaded by developmental genetic investigations of the spiders *Cupiennius salei* and *Parasteatoda tepidariorum* (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2010; Damen et al., 1998; Schwager et al., 2007; Schwager et al., 2017; Stollewerk et al., 2003). Such works have investigated several important aspects of chelicerate development, such as the molecular mechanism of dorsoventral axis patterning (Akiyama-Oda and Oda, 2003; Akiyama-Oda and Oda, 2006; Oda and Akiyama-Oda, 2008; Pechmann et al., 2017), segmentation of the prosoma (Damen et al., 1998; Paese et al., 2018; Pechmann et al., 2017; Schwager et al., 2009; Setton and Sharma, 2018) and opisthosoma (Paese et al., 2018; Stollewerk et al., 2003), and specification of prosomal versus opisthosomal fate (Khadjeh et al., 2012; Pechmann et al., 2015). A few non-spider chelicerate models have also emerged in recent years and provided new perspectives on chelicerate development, such as the horseshoe crab *Limulus polyphemus* (Xiphosura) (Blackburn et al., 2008; Mittmann and Scholtz, 2001), the mites *Archegozetes longisetosus* (Barnett and Thomas, 2012; Barnett and Thomas, 2013a; Barnett and Thomas, 2013b; Telford and Thomas, 1998) and *Tetranychus urticae* (Acariformes) (Dearden, 2002; Grbić et al., 2007; Khila and Grbić, 2007), the tick *Rhipicephalus microplus* (Parasitiformes) (Santos et al., 2013), the Arizona bark scorpion *Centruroides sculpturatus* (Scorpiones) (Sharma et al., 2014a) and the harvestmen *Phalangium opilio* (Opiliones) (Sharma et al., 2012a; Sharma et al., 2012b; Sharma et al., 2013; Sharma et al., 2015a). Studies in *A. longisetosus* and *C. sculpturatus* have suggested that changes in Hox gene number and spatial expression are responsible for such phenomena as the reduced segmentation of mites and the supernumerary posterior appendage identities of scorpions (Barnett and Thomas, 2013a; Sharma et al., 2014a).

In addition to the insights into chelicerate development, one of the major recent outcomes of

increasing availability of genomic resources in the group was the discovery of a shared whole (or partial) genome duplication in spiders and scorpions. This genome duplication is inferred to trace back to the most recent common ancestor of the recently proposed clade Arachnopulmonata, which includes spiders, scorpions, and three other arachnid orders with book lungs (Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2014a; Sharma et al., 2014b; Sharma et al., 2015b). This duplication event is also inferred to be independent from the multiple rounds of whole genome duplication undergone by horseshoe crabs (Xiphosura) (Kenny et al., 2015; Shingate et al., 2020a). The duplication in spiders and scorpions encompasses several important developmental genes, such as those in the homeobox family (Leite et al., 2018; Schwager et al., 2017). Excitingly, there is increasing evidence of divergent expression and function of paralogs, such as in the case of Hox genes, leg gap genes and retinal determination gene network (RDGN) homologs (Gainett et al., 2020; Samadi et al., 2015; Schomburg et al., 2020; Schwager et al., 2017; Sharma et al., 2014a; Turetzek et al., 2015; Turetzek et al., 2017).

Nevertheless, knowledge on the development of several arachnid orders still remains scarce or entirely unexplored in the contexts of embryological study and genomic architecture. One particularly interesting arachnid order whose genomic evolution and developmental biology remains largely unexplored is Amblypygi. Commonly known as whip spiders, Amblypygi comprise approximately 220 described species of nocturnal predators that are distributed primarily in the tropics and subtropics worldwide (Miranda et al., 2018; Weygoldt, 2000). Amblypygi is part of the clade Pedipalpi (together with the orders Thelyphonida and Schizomida), which in turn is sister group to spiders (Araneae) (Ballesteros and Sharma, 2019; Giribet, 2018; Lozano-Fernandez et al., 2019; Sharma et al., 2014b). Considering the phylogenetic position of Amblypygi within Arachnopulmonata, a better understanding of Amblypygi genomics is anticipated to facilitate exploration of the evolutionary outcomes of

the genome duplication in Arachnopulmonata, and better characterizing the extent to which arachnopulmonate orders retained and specialized the ensuing paralogous genes. Unfortunately, the embryology of the order Amblypygi is known only from the works of Pereyaslawzewa (Pereyaslawzewa, 1901), a brief mention in Strubell (Heymons, 1904) ("Phrynidae"), Gough (Gough, 1902), and the comprehensive study of Weygoldt (1975) in *Phrynus marginemaculatus* (formerly *Tarantula marginemaculata*).

Beyond the interest in their genomic architecture, whip spiders possess a fascinating biology and natural history. Whip spiders are emerging as model systems in behavioral ecology, learning, and neurophysiology in arachnids (Chapin and Hebets, 2016; Wiegmann et al., 2016). Their pedipalps, the second pair of appendages, are robust raptorial devices used for striking prey, including mostly arthropods, and even small vertebrates (Chapin and Hebets, 2016; Seiter et al., 2019; Weygoldt, 2000). However, their most conspicuous characteristic and namesake is their "whip", a modified antenniform first walking leg that is not used for locomotion, but instead as a sensory appendage. The antenniform legs are extremely elongated relative to the other leg pairs, and the tibia and tarsus are pseudo-segmented into hundreds of articles that harbor an array of chemo-, thermo-, hygro- and mechanoreceptive sensilla (Foelix et al., 1975; Igelmund, 1987; Santer and Hebets, 2011b). The peripheral circuitry of the antenniform legs is complex, exhibiting peripheral synapses and giant neurons which are possibly involved in fast sensory responses (Foelix, 1975; Foelix et al., 2002; Santer and Hebets, 2011b). The antenniform legs also have been shown to be important for foraging, mating, and intrasexual contests (Santer and Hebets, 2008; Santer and Hebets, 2011a). Notably, concentration of sensory structures, elongation and pseudo-segmentation of these legs constitute a striking convergence with the antennae of mandibulate arthropods (Pancrustacea and Myriapoda). Evidence from Hox gene expression and functional experiments support the view that the antenna of mandibulates is positionally homologous to the chelicera of chelicerates

(deutocerebral appendage) (Sharma et al., 2015a; Telford and Thomas, 1998). Despite their different positions along the antero-posterior axis of the body, the serial homology of both antennae (mandibulates) and antenniform legs (whip spiders) to walking legs constitutes a potentially useful comparison to address whether a striking morphological convergence is associated with convergence in mechanisms of genetic patterning.

Much of our knowledge about the genetic basis for appendage fate specification in arthropods was dis- covered in the fruit fly *Drosophila melanogaster* and involves regulation by leg gap genes and Hox genes. Broadly overlapping domains of leg gap genes *homothorax* (*hth*), *dachshund* (*dac*) and *Distal-less* (*Dll*) in the antenna promote antennal fate through activation of downstream antennal determinants such as *spineless* (Casares and Mann, 1998; Dong et al., 2001; Dong et al., 2002; Duncan et al., 1998) (reviewed by [Setton et al., 2017]); leg identity in thoracic appendages is mediated by mutually antagonistic *hth* (proximal), *dac* (median) and *Dll* (distal) domains, and repression of *hth* by the Hox gene *Antennapedia* (*Antp*) (Casares and Mann, 1998; Struhl, 1981; Struhl, 1982). Ectopic expression of *hth* on the leg discs results in leg-to-antenna transformation, whereas ectopic *Antp* in the antennal disc results in antenna-to-leg transformation (Casares and Mann, 1998; Struhl, 1981; Struhl, 1982). In chelicerates, similar to *D. melanogaster*, the chelicera (antennal positional homolog) has broadly overlapping domains of *Dll* and *hth*; legs and pedipalps lack a distal *hth*, and a proximal *Dll* expression domain, indicating a conservation of expression of positional homologs (Pechmann and Prpic, 2009; Prpic and Damen, 2004; Prpic et al., 2003; Sharma et al., 2012a). In line with this evidence, knockdown of *hth* in the harvestmen *Phalangium opilio* results in chelicera-to-leg transformation (and also partial pedipalp-to-leg transformation) (Sharma et al., 2015a). In contrast to *D. melanogaster* and other insects, arachnid *Antp* is not expressed in the leg-bearing segments, and knockdown of this gene in the spider *Parasteatoda tepidariorum* has revealed a role in repressing limbs in the opisthosoma (Khadjeh et al., 2012). Expression of *dac* in a medial

domain of chelicerate pedipalps and legs suggests a conserved role in patterning the medial leg territory; accordingly, knockdown of *dac* in the harvestmen *P. opilio* and the spider *P. tepidariorum* results in defects in medial podomeres, namely, the femur, patella, and tibia (Sharma et al., 2013; Turetzek et al., 2015). Additionally, loss of a *dac* domain of the chelicera of spiders in comparison to harvestman has been implicated in the transition from three- to two-segmented chelicerae (Sharma et al., 2012a; Sharma et al., 2013).

Towards revitalizing embryological studies in Amblypygi, we generated herein protocols and comprehensive genomic resources for developmental genetic study of the whip spider *P. marginemaculatus* (Fig. 1a, b), resuming the seminal studies begun by Peter Weygoldt with this promising model species. Leveraging a comprehensive developmental transcriptome we established for this species, we first investigated the presence and retention of systemic gene duplications in Amblypygi predicted from their phylogenetic position in Arachnopulmonata; to mitigate biases stemming from survey of a data point from a single exemplar, we included as well the embryonic transcriptomes we recently generated for two species of the distantly related family Charinidae (*Charinus ioanniticus* and *Charinus israelensis*) (Gainett et al., 2020; Garwood et al., 2017; Weygoldt, 2000). As a first step towards elucidating the patterning of antenniform legs, we investigated a possible role of the leg gap genes *homothorax*, *dachshund* and *Distal-less* in specifying antenniform leg fate, by characterizing their expression pattern in developing embryonic stages of *P. marginemaculatus* using newly established procedures for colorimetric whole mount in situ hybridization.

Results

Overview of embryogenesis in Amblypygi

A characterization of the postembryonic development of *P. marginemaculatus* is available in

the works of Weygoldt (Weygoldt, 1969; Weygoldt, 1970). The following summary is after Weygoldt (1975), complemented with observations from the present study. We note that the timing of embryonic stages is approximate, as the temperature in a previous work (Weygoldt, 1975) was kept in a range of 24–27 °C. Details on the culture and collection of embryos are described in “Methods”.

In our colony, a sample of 18 clutches from different females had an average of 19 eggs (minimum: 8 eggs; maximum: 39 eggs). Each egg is ovoid and measures 1.2–1.4 mm at the major axis. The yolk granules are light green, and fixation turns them light yellow. Upon secreting the brood sac and eggs into the ventral side of the opisthosoma, the brood sac is transparent, but gradually hardens and becomes dark brown in the course of two days. Around six days after egg laying (6 dAEL), a blastoderm is formed. At 7–8 dAEL, the yolk is more condensed, a perivitelline space is formed, and cells move into the yolk presumably in the region of the future blastopore. At 8–9 dAEL, a blastopore is recognizable on the dorsal surface of the embryo, and a germ disc is formed. Towards 10 dAEL, the germ band starts to take form, the blastopore region is displaced posteriorly, and the first signs of segmentation of the prosoma are visible. A presumable cumulus, a group of migrating cells with dorsoventral organizing properties, was reported by Weygoldt (1975) at this stage, but was not investigated by us. At 11–12 dAEL, the nascent prosomal segments elongate anteriorly and by 13–14 dAEL the segmented germ band occupies the ventral hemisphere of the egg. The limb buds of prosomal appendages begin to appear (Fig. 2a). The process of reversion begins, with a sagittal split of the germ band (Fig. 2a). From this point, until 19 dAEL, the ventral sulcus progressively enlarges with the dorsal displacement of both halves, and then begins to ventrally fuse again starting anteriorly (Fig. 2a–e). The head lobes on 15 dAEL have a circular outline and uniform ectoderm (Fig. 2f). From 15–19 dAEL, ectodermal cells invaginate along the ventral midline and head lobes (neural precursor cells), which can be seen as evenly spaced dark dots (Fig. 2f–

h). Concomitantly, the anterior rim of the head lobes forms semi-lunar grooves (i.e., folds over itself) (Fig. 2f–h). Opisthosomal segments 1–12 sequentially appear from the posterior growth zone (Fig. 2i–k). At 19 dAEL, the yolk moves into the caudal papilla (growth zone) and the opisthosoma flexes ventrally (Fig. 2k, n). The first pair of legs (future antenniform legs) is already longer than the other appendages in an early limb bud stage (15 dAEL), and more clearly in subsequent stages (Fig. 2l–n). A protuberance on the proximal end of the second walking leg pair (L2), the nascent lateral organ (osmoregulatory embryonic organ), is discernible at 16 dAEL and later stages, and becomes reniform (Fig. 2l–n). From 20–21 dAEL, the embryo begins to secrete a cuticle, which is darker on the fully formed lateral organ. The head lobes have completely folded onto themselves, and the distal segment of the chelicera is sclerotized. The embryo then ruptures the egg shell, and is now termed deutembryo (Fig. 2o–q). Deutembryos remain in the brood sac for further 70 days without discernible external changes. The primordia of the eyes are formed towards the end of deutembryo development (70–106 dAEL), and eye spots of the median and lateral eyes become visible. From 90–106 dAEL the deutembryos molt into praenymphae, leave the brood sac and climb onto the adult female's back (Fig. 1b). For a histological description of gastrulation, of the mesoderm, brain and nervous system, digestive system and gonad formation, we refer the reader to the exquisite original descriptions of Weygoldt (1975).

Phrynus marginemaculatus embryonic transcriptome

In order to provide a comprehensive resource for facilitating developmental genetic study and comparative genomics in Amblypygi, we assembled an embryonic transcriptome spanning different embryonic stages. The assembly of *P. marginemaculatus* transcriptome resulted in 544,372 transcripts composed of 277,432,101 bp and N50 of 440 bp. BUSCO scores, which reflect the presence and number of copies of widely conserved single- copy genes (BUSCO-

Arthropoda; *n* = 1066), indicated 94.2% completeness with 5.9% of genes duplicated. From all transcripts (including isoforms), we predicted 37,540 peptide sequences with the Trinotate pipeline with a top blastp hit in Swiss-Prot database. A total of 801 of these peptides (2.1%) had a top hit to non-metazoan proteins, which we presumed to derive from microbial contaminants. The total number of predicted proteins is higher in comparison with the predicted protein set from the genome of the spider *Parasteatoda tepidariorum* (27,900) and the scorpion *Centruroides sculpturatus* (30,456) (Schwager et al., 2017), which is attributable to the typical fragmentation of genes in such transcriptomic assemblies.

To investigate the possibility of systemic gene duplication in Amblypygi and further test the transcriptome completeness, we searched for all ten Hox genes, emphasizing discovery of paralogs known from the genomes of the spider *P. tepidariorum* (except for *fushi tarazu*) and the scorpion *C. sculpturatus* (except for *Hox3*) (Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2014a). One of the *Hox3* copies of *P. tepidariorum* has a highly divergent homeodomain sequence, constituting an unusual long branch in our phylogenetic analysis (Additional file 1: Fig. S1). The same is true for the *Hox3* paralog *bicoid* of *Drosophila melanogaster* (Additional file 1: Fig. S1). We therefore removed both paralogs from our main analysis. We found orthologs of all ten Hox genes in *P. marginemaculatus*, namely *labial*, *proboscipedia*, *Hox3*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultra-bithorax*, *abdominal-A* and *Abdominal-B* (Fig. 3a, b). We found evidence that all Hox genes, except *proboscipedia* exhibited two copies, and the predicted protein sequences of these copies overlapped with clear amino acid differences in *P. marginemaculatus* transcriptome (Fig. 3b). A *Pmar-Dfd* and *Pmar-abdA* were assembled in a third gene, but their mostly non-overlapping sequences on the alignment and nearly identical nucleotide or amino acid sequence to their respective paralog 2 indicate that they are fragmented assemblies of the same copy. We additionally analyzed two embryonic transcriptomes of *Charinus* whip spiders recently

generated by us (Gainett et al., 2020). For *C. israelensis*, we discovered orthologs of all ten Hox genes, with two genes for all cases except *ftz* and *Antp* (Fig. 3a, b). A third gene was annotated as *Cisr-Dfd*, but we interpret it as an assembly error, given its largely non-overlapping sequence with respect to the other *Cisr-Dfd* sequences. For *C. ioanniticus*, we discovered one gene each for *ftz*, *Antp*, *Ubx*, *AbdB*, and two genes each for *Hox3*, *Dfd*, and *Scr*. We did not detect orthologs of *lab*, *pb* and *AbdB* in *C. ioanniticus* transcriptome (Fig. 3a, b). These absences almost certainly reflect incompleteness of transcriptomic assemblies, and cannot be equated with true losses. We therefore consider the union of the Hox gene counts across the three species as the reconstruction of the whip spider ancestral Hox complement. Our results suggest that the ancestral Hox complement in Amblypygi consists of 20 genes (two copies of every Hox gene), a trait they share with spiders and scorpions. While a genome is not available for whip spiders, the phylogenetic placement of this order, as well as the pattern of Hox duplication, strongly suggests that whip spiders share the condition of two Hox clusters observed in spider and scorpion genomes.

Gene expression analyses in *Phrynus marginemaculatus*

All three methods we trialed for fixing embryos of *P. marginemaculatus* yielded embryos which are suitable for in situ hybridization (ISH). However, fixation of intact embryos in formaldehyde/heptane phase sometimes results in the vitelline membrane adhering to the embryo. Upon dissection of the vitelline membrane for ISH, parts of the embryo are commonly lost. Piercing a hole in the embryo for fixation in formaldehyde achieves a similar result, and is therefore not recommended. Dissecting the vitelline membrane under PBS and immediately fixing the embryos in formaldehyde yielded the best results. We did not detect any difference in the signal-to-background ratio of colorimetric ISH between embryos fixed for 2 h or 24 h. To test gene expression assays and investigate the genetic patterning of antenniform legs, we

selected the leg gap genes *Distal-less*, *dachshund* and *homothorax*, which constitute the subset of PD axis patterning genes associated with functional (RNAi) datasets in other arachnid model species (Schoppmeier and Damen, 2001; Sharma et al., 2013; Sharma et al., 2015a).

Distal-less identification and expression

Similar to other arachnopulmonates and arthropods in general, a single *Distal-less* gene, *Pmar-Dll*, was found by similarity searches and its orthology to other metazoan *Distal-less* genes was confirmed with a phylogenetic analysis (Additional file 2: Fig. S2). *Pmar-Dll* is strongly expressed in all developing appendages, from early limb bud stage (15 dAEL) and is localized to the distal portion of the chelicera, pedipalp and legs throughout leg elongation (Figs. 4a–i; 5a–l). In 17–18 dAEL embryos, the expression is heterogeneous and forms bands of stronger expression on the forming podomeres of the pedipalps and legs (Fig. 5h–l). We did not detect expression on the ventral midline (*contra* [Mittmann and Scholtz, 2001; Sharma et al., 2012a]), or in the opisthosoma, and observed only diffuse expression on the head lobes of the stages surveyed. Expression on L1 of 16–17 dAEL embryos is subtly stronger on the proximal part of the expression domain (Fig. 5c–c′), but we did not detect clear differences in younger or older embryonic stages. The proximal and distal boundaries of strong *Distal-less* expression do not differ between L1 and the other legs, spanning two-thirds of the legs from the tip of the tarsus and excluding the most proximal third (Fig. 5c, i).

In contrast to other arachnid embryos (mygalomorph spiders [Pechmann and Prpic, 2009]; araneomorph spiders [Prpic and Damen, 2004; Schoppmeier and Damen, 2001]; Opiliones [Sharma et al., 2012a]; and mites [Barnett and Thomas, 2013b; Thomas and Telford, 1999]), *Dll* expression was not detected as a strong domain in the endite of pedipalp, which is very inconspicuous at these embryonic stages. We were not able to assess *Dll* expression in the pedipalpal endite in deutembryo stages due to deposition of cuticle.

dachshund identification and expression

dachshund is typically present as a single copy in Arthropoda, but is found independently duplicated in Arachnopulmonata and Xiphosura (Nolan et al., 2020; Turetzek et al., 2015). We found two genes with high similarity to *Dmel-dac* and confirmed their identity as *dachshund* paralogs with a phylogenetic analysis (Additional file 3: Fig. S3). *Pmar-dac1* and *Pmar-dac2* are each nested in a clade with paralogs 1 and 2 of other arachnopulmonates, which appears as independent from horseshoe crab duplications (Additional file 3: Fig. S3).

Pmar-dac1 is strongly expressed as broad medial band in the pedipalp and the four pairs of legs (Fig. 6a–f). This sharp medial band excludes the most proximal and the distal domains of these appendages (Fig. 7). No staining was detected on the chelicera (Figs. 6a–f, 7a, e). Additionally, expression occurs in three domains in the head lobes and strongly on the ventral midline of the prosoma. The continuous expression on the ventral midline stops before the opisthosoma (Fig. 6a–f).

Pmar-dac2 is expressed as ring at the proximal region of chelicera, pedipalp and legs (Fig. 6g–i). On 16–17 dAEL embryos, a second thinner ring of expression is also present on the pedipalps and legs (Figs. 6g–i, 7i–l). Expression domains also occur on the neuromere of L1 and pedipalp, being larger on the pedipalpal neuromere (Fig. 6g). Three spot domains also occur on O1, O2 and O3 segments (Fig. 6h–i).

homothorax identification and expression

Similar to the case with arachnopulmonate *dachshund*, *homothorax* is also found duplicated in arachnopulmonates (Leite et al., 2018; Turetzek et al., 2017). We annotated two genes in *P. marginemaculatus* transcriptome and the phylogenetic analysis shows that each copy is nested

in two clades with the arachnopulmonate paralogs (Additional file 4: Fig. S4). *Pmar-hth1* is conspicuously expressed on the developing appendages throughout the stages studied (15–18 dAEL; Figs. 8, 9). On 16–17 dAEL embryos, *Pmar-hth1* is expressed more strongly on the proximal parts of the chelicera, pedipalp, and legs, and is absent in a small distal domain (Fig. 9a–d). On 17–18 dAEL, expression in all elongated appendages is similar, and spans most of the appendage length except for a small distal territory (Fig. 9e–h). Additionally, three stronger stripe domains occur distal to the coxa of pedipalp and legs. Expression in L1 and other legs is similar in the studied stages, save for a slightly expanded distal domain on the proximal tarsus of L1 (Fig. 9g, h). Expression on the ventral ectoderm and head lobes is weak and homogenous in early stages, and becomes stronger in 17–18 dAEL (not shown) *Pmar-hth2* is expressed in the developing appendages, but is reduced in intensity and different in pattern in com- parison to *Pmar-hth1* (Fig. 8b–b′). *Pmar-hth2* spans the whole chelicera (Fig. 9i). Expression on the pedipalp spans most of the appendage, excluding only a distal territory (Fig. 9j). In addition, there is a proximal and a sub-terminal strong domain (Fig. 9j). Expression on L1 and other legs is identical in all stages studies. *Pmar-hth2* expression on the legs is consistently weaker in comparison to the pedipalp, but likewise possess a proximal and a distal sub-terminal domain (Fig. 9k, l). Expression on the ventral ectoderm, opisthosoma and head lobes is ubiquitous (Fig. 8b).

Discussion

Genomic resources for whip spiders corroborate widespread retention of duplicates across Arachnopulmonata

Changes in Hox gene expression, regulation and targets are linked to changes in morphology and Hox gene duplications have also been implicated in morphological innovations (Hughes

and Kaufman, 2002; Hughes CL and Kaufman TC, 2002; Ronshaugen et al., 2002; Wagner et al., 2003). Duplication in the Hox cluster is common in the vertebrates, which may have up to eight Hox clusters (teleost fishes) due to successive rounds of genome duplications (Wagner et al., 2003). On the other hand, genome and Hox cluster duplications are considered rare in Arthropoda (Pace et al., 2016). Complete Hox cluster duplication in the phylum has only been reported in horseshoe crabs (Xiphosura) (Kenny et al., 2015; Nossa et al., 2014; Shingate et al., 2020b; Zhou et al., 2020), and in spiders and scorpions (presumed ancestral in Arachnopulmonata). The phylogenetic placement of Amblypygi makes this group an opportune system to test the downstream prediction that *P. marginemaculatus* retains and expresses Hox paralogs shared by spiders and scorpions. In accordance with this hypothesis, we discovered that the whip spider

P. marginemaculatus presents homologs of all ten Hox genes, and all excepting *proboscipedia* appear to be duplicated, as supported by our alignment and phylogenetic analysis. Our survey in the two *Charinus* transcriptomes confirmed the duplication of several Hox genes, and additionally revealed that two copies of *proboscipedia* (absent from *P. marginemaculatus transcriptome*) are present at least in *C. israelensis*. Our results suggest that Amblypygi retain a complete duplicated Hox complement with 20 genes. Similarly, a recent transcriptomic survey in *Charinus acosta* and *Euphrynichus bacillifer* reported duplications of nearly all Hox genes (excepting *proboscipedia*), as well as *Wnt* and *Frizzled* genes (Harper et al., 2020). Taken together with the recent discovery of duplicated RDGN genes in *Charinus*, our results suggest that Amblypygi retain ohnologs of numerous developmental patterning genes, as a function of an ancient, shared whole genome duplication.

A second line of evidence of a shared genome duplication event comes from conserved expression patterns of leg gap gene paralogs. It has recently been shown that duplicates of leg patterning genes *extradenticle*, *optomotor blind*, *dachshund* (*dac*) and *homothorax* (*hth*)

present one copy with an expression domain resembling a stereotypical panarthropod pattern, and a second copy with a divergent or novel expression pattern, with these two sets of expression patterns shared by a spider and a scorpion (Nolan et al., 2020). For instance, *dac1* of surveyed arachnopulmonates retains the plesiomorphic expression pattern of a broad median band in the appendages, the condition observed in the single- copy *dac* of all non-arachnopulmonate arachnids surveyed to date (Barnett and Thomas, 2013b; Sharma et al., 2012a), whereas *dac2* is expressed as a proximal and a separate medial domain at the patella-tibia boundary (Nolan et al., 2020; Pechmann and Prpic, 2009; Prpic and Damen, 2004; Turetzek et al., 2015). Similarly, *hth1* retains a plesiomorphic expression pattern on the proximal-median appendage excluding the tarsus, whereas *hth2* presents a strong short proximal domain and stripes of expression at podomere boundaries (Nolan et al., 2020; Pechmann and Prpic, 2009; Prpic and Damen, 2004; Turetzek et al., 2015). In the whip spider *P. marginemaculatus*, we indeed recovered two copies of *dachshund* and *homothorax*, with paralogs nested in two clades of arachnopulmonate copies with high nodal support (Additional file 3: Fig. S3; Additional file 4: Fig. S4). Moreover, *Pmar-dac* and *Pmar-hth* have paralog-specific expression patterns shared with spiders and a scorpion: *Pmar-dac1/2* and *Pmar-hth1/2* conserve the plesiomorphic expression pattern of arachnids, while copy 2 retains the arachnopulmonate-specific pattern. These data, together with the evidence from Hox gene duplications and the topologies of the *dac* and *hth* gene trees, provide compelling support for genome duplication as a synapomorphy of Arachnopulmonata.

A *dachshund2* domain in the two-segmented clasp-knife chelicera of *P. marginemaculatus*

Expression of *Pmar-Dll*, *Pmar-dac1/2* and *Pmar- hth1/2* on the appendages is largely similar to spiders and scorpions (see above) (Nolan et al., 2020; Pechmann and Prpic, 2009; Prpic and Damen, 2004; Prpic et al., 2003; Turetzek et al., 2015). A notable difference in the chelicera

of the whip spider is the retention of proximal expression of one of the *dachshund* paralogs. In arachnids with three-segmented chelicerae, such as the daddy-long-legs *Phalangium opilio*, *dac* is expressed in the proximal segment of the appendage (Sharma et al., 2012a). The same is true for the three-segmented chelicera of scorpions, in which both *dac1* and *dac2* are expressed in the proximal segment (Nolan et al., 2020). In the two-segmented chelicera of spiders, *dac1* is not expressed in the chelicera save for a non-ectodermal faint staining in late embryonic stages (Abzhanov and Kaufman, 2000; Pechmann and Prpic, 2009; Prpic and Damen, 2004; Prpic et al., 2003; Turetzek et al., 2015). *dac2* of spiders is also largely not expressed in the chelicera, save for a small ventral domain at the base of the chelicera of *Parasteatoda tepidariorum* and *Pholcus phalangioides* (Turetzek et al., 2015). It has been proposed that the loss of a proximal *dac* domain is associated with the loss of the proximal segment in the chelicera of spider and other arachnids. This hypothesis is supported by functional data in the daddy-long-leg *P. opilio*: knockdown of the single-copy *dac* resulted in the loss of the proximal segment of the chelicera (as well as deletions of medial podomeres in pedipalps and legs) (Sharma et al., 2013). Similar to spiders, the two-segmented chelicera of the whip spider *P. marginemaculatus* does not express *Pmar-dac1*. Nonetheless, a clear complete ring of proximal expression was detected for *Pmar-dac2*.

In contrast to the classic leg gap phenotype incurred by knockdown of the ancestral *dac* copy in the harvest- man, knockdown of *dac2* in the spider *P. tepidari*orum resulted in the loss of the patella-tibia segment boundary in the pedipalp and walking legs; effects of this knock-down on the chelicera were not reported by the authors, nor is the function of *dac1* known in *P. tepidariorum* (Turetzek et al., 2015). Given the interpretation that *dac2* has a derived role in patterning a segment boundary in the spider, the incidence of the *Pmar-dac2* stripe in the middle of the whip spider chelicera (and the absence of *Pmar-dac1* in the chelicera) is tentatively hypothesized here to constitute evidence of a segment boundary patterning activity

that separates the basal and distal cheliceral segment of Amblypygi (rather than evidence of a proximal segment identity in whip spiders that is homologous to the proximal cheliceral segment of groups like harvestmen and scorpions). Exploring this hypothesis in future will require the establishment of RNA interference tools in *P. marginemaculatus.*

Antenniform leg distal allometry is specified early in development

To investigate a possible role of changes in leg gap gene expression in the patterning of the antenniform leg pair (L1) of whip spiders, we specifically searched for changes in expression domain between embryonic antenniform leg (L1) and the posterior legs (L2–4). The expression patterns of *Pmar-Dll*, *Pmar-dac1/2* and *Pmar-hth1/2* are mostly identical between the antenniform leg L1 and the posterior legs. We found only a slightly expanded *Pmar- hth1* domain in the distal leg (compare Fig. 9g, h). It is unclear if this difference bears any functional significance in the differential fate of L1. Similar distal expression dynamics appear to distinguish the identity of the pedipalp and walking legs of the harvestman, as knockdown of the single copy of *hth* in *P. opilio* results in homeotic transformation of pedipalps to legs (Sharma et al., 2015a).

Besides the difference in sensory equipment, the antenniform leg pair of adults differ from the posterior walking legs in being much longer, and presenting numerous articles (pseudo-segments) on the tibia, metatarsus and especially the tarsus, whereas the walking legs have pseudo-segments only in the tarsus (Igelmund, 1987; Santer and Hebets, 2011b; Weygoldt, 2000). Antenniform legs may be more than 2.5 times longer than the walking legs, and this allometry in adults is mostly accounted for by the length added by the distal leg segments, which may attain more than 100 articles (tibiomeres + tarsomeres) (Igelmund, 1987)Understanding the timing of the ontogenetic divergence of L1 relative to the walking legs is crucial for circumscribing target stages for investigation of the genetic patterning of this

phenomenon.

Current evidence indicates that this allometry on the antenniform legs takes place both embryonically and postembryonically. The praenympha, the stage that molts from the deutembryo and climbs to the mother's back, has the full number of tibiomeres, but not of tarsomeres. The final number of tarsomeres is attained in the next molt to the protonymph (Beck et al., 1977; Igelmund, 1987), and from this moment onwards the subsequent growth is isometric (Beck et al., 1977; Igelmund, 1987). In embryos of *P. marginemaculatus*, Weygoldt (Weygoldt, 1975) observed that the primordia of L1 are the first to appear in the germ band and are already longer than the remaining prosomal appendages, an observation paralleling the allometric growth of the L2 limb bud of the harvestman *P. opilio* (wherein L2 is the longest leg and also bears the most tarsomeres) (Sharma et al., 2012b). It is presently unclear at what point in development the distal podomeres of antenniform legs attain their allometry with respect to the proximal podomeres in the whip spider. In other words, are embryonic legs simply growing longer overall with respect to walking legs, or are they longer due to distal growth? The expression pattern *dachshund* in *P. marginemaculatus* early embryos reveals a further interesting aspect of this allometry. The expression of *dac* (including paralogs) in the legs of arachnids surveyed to date occurs on podomeres proximal to the tibia, so that the tibia–metatarsus–tarsus territory may be identified as the a *dac*-free domain (Abzhanov and Kaufman, 2000; Barnett and Thomas, 2013b; Nolan et al., 2020; Pechmann and Prpic, 2009; Prpic and Damen, 2004; Sharma et al., 2013; Turetzek et al., 2015). The distal boundary of *Pmar-dac1* and *Pmar-dac2* in stages long before podomere formation reveals that the distal domain (*Pmar-dac1*-free cells) of L1 is already longer with respect to the proximal leg and the other legs (com pare Fig. 7c, d). These data support the inference that the distal allometry, and by extension the antenniform leg fate, are already specified from the onset of limb bud development.

Conclusion

The study of the embryology of Amblypygi has faced 45 years of hiatus since the seminal work describing the development of *Phrynus marginemaculatus* (Weygoldt, 1975). Here, we take the first steps in establishing *P. marginemaculatus* as a chelicerate model for modern evolutionary developmental study and for comparative genomics. We provide evidence for a genome duplication in Amblypygi that is shared with the remaining arachnopulmonate orders; show that the distal allometry of antenniform legs is specified early in appendage embryogenesis; and observe that expression patterns of the leg gap genes *Distal-less*, *dachshund* and *homothorax* do not substantiate their role in fate specification of the sensory legs of whip spiders. Future efforts to uncover the genetic underpinnings of this fascinating convergence with the mandibulate antenna should prioritize discovery of leg fate specification mechanisms in early whip spider development. Despite the establishment of a few promising arachnid models for studying chelicerate embryology, a pervasive challenge has been the establishment of toolkits for testing gene function (Akiyama-Oda and Oda, 2003; Khila and Grbić, 2007; Schoppmeier and Damen, 2001; Sharma et al., 2013). Therefore, future efforts to study whip spider development must emphasize the development of tools for functional experiments in *P. marginemaculatus*.

Methods

Establishment of amblypygid colony in the laboratory

Adult individuals of *Phrynus marginemaculatus* were purchased from a commercial supplier

that collects animals in the Florida Keys (USA). The maintenance of a colony of *P. marginemaculatus* (Fig. 1a, b) in the laboratory followed guidelines provided in the works of Peter Weygoldt (Weygoldt, 1969; Weygoldt, 1970; Weygoldt, 1975) and notes on behavioral studies of this species (Fowler-Finn and Hebets, 2006; Rayor and Taylor, 2006; Santer and Hebets, 2009). While most species of Amblypygi have been reported as aggressive and better kept in separate containers (Weygoldt, 2000), individuals of *P. marginemaculatus* can be kept in a stable laboratory colony in large community containers (Rayor and Taylor, 2006). A maximum of 20 animals may be allowed to interact and mate freely in plastic boxes (67.3 × 41.9 × 32.4 cm) with a 2 cm layer of damp coco- nut fiber saturated with distilled water. The substrate is continuously dampened with approximately 500 mL of dH_2O on the substrate every week. All animals are kept in an acclimatized room at 26–28 °C in a controlled reversed 16L:8D light cycle to simulate summer months in Florida. To prevent flies from reproducing in this humid environment, the lid of the container is provided with a 10 × 20 cm fine mesh cloth, which also allows for air circulation. Each container is furnished with pieces of tree bark, egg carton, and retreats made from polystyrene foam or florist's foam, to provide enough space for animals to hide and suspend themselves during molting and egg-laying. After about 2 weeks of setting up a terrarium, the substrate hosted a stable colony of collembolans (possibly brought with the tree bark). The co-occurrence of collembolans does not appear to be disadvantageous in any way. Collembolans are invariably observed feasting on carcasses and likely contribute to preventing mite infestations (see also [Weygoldt, 2000]). Animals are fed at least twice a week with nymphs of *Acheta domestica ad libitum*. Carcasses are removed as soon as detected to prevent the spread of mites. Freshly hatched praenymphae are kept with the female, and aggregate on the back of the brooding female (Fig. 1b). The next instar, the protonymph, disperse from the female upon molting. Protonymphs are transferred to community terraria as described above. Protonymphs and subsequent instars are fed pinheads of *A. domestica* and

Gryllus bimaculatus ad libitum three times per week until adulthood. The first adult individuals from our laboratory-bred animals were detected 8 months after hatching, which accords to what has been described by Weygoldt (1975).

Terraria with adults are inspected daily for freshly molten individuals or ovigerous females, which are immediately separated from the main colony to avoid aggressive interactions. Ovigerous females carry the eggs outside the body in a brood sac attached to the ventral side of the opisthosoma (Weygoldt, 2000). Eggs are allowed to develop with the female until the desired stage for developmental work. Approximate stages may be inferred from days after egg laying (dAEL) at 26 °C following the staging in (Weygoldt, 1975). For obtaining the eggs, females are anesthetized with CO_2 and immobilized in customized plastic petri dishes (35 × 11 mm) with a central whole and foam (Fig. 10a, b). With a blunt forceps pair, the pleura of the opisthosoma embracing the eggs may be gently displaced from the brood sac, and the brood sac removed as whole (Fig. 10c). The female may be returned to the colony unharmed. The brood sac easily dissolves when immersed in a phosphate buffer solution (PBS), and eggs may be cleaned with forceps under PBS. We obtained high mortality in keeping isolated eggs developing in petri dishes with humid (wet filter paper) or dry in 26 °C incubators, as mold frequently develops. However, we found out that embryos will survive and develop immersed in PBS supplemented with kanamycin (1:5000) at 26 °C incubator and even hatch into deutembryos (~ 20 dAEL).

Phrynus marginemaculatus transcriptome assembly

We collected embryos of different embryonic stages as described above, and conducted total RNA extractions of each stage separately: hatched deutembryo (female #5, $n = 3$; undetermined age, before eyespots), unhatched deutembryo (female #2 $n = 3$; approx. 21 dAEL), and leg elongation stage (female #3, $n = 2$; approx. 16–17 dAEL), late limb bud stage (female 6, $n =$

1; 16 dAEL) and early limb bud stage (female 6, *n* = 1; 14 dAEL) (voucher images are available upon request). Embryos were homogenized in Trizol TRIreagent (Invitrogen) and stored at − 80 °C until extraction. The combined total RNA of the five extractions was submitted to cDNA library preparation and sequencing in an Illumina High-Seq 2500 platform using paired-end short read sequencing strategy (2 × 100 bp) at the Center for Gene Expression, Biotechnology Center of University of Wisconsin-Madison (USA). The quality of the raw reads was assessed with FastQC (Babraham Bioinformatics). De novo assembly of compiled paired-end reads was conducted with the software Trinity v. 2.6.6 (Grabherr et al., 2011), using Trimmomatic v. 0.36 (Bolger et al., 2014)to remove low quality reads and adaptors. Completeness of the transcriptome was assessed with the Trinity script '*TrinityStats.pl*', and BUSCO v. 3 (Waterhouse et al., 2017) on the longest isoforms using the database 'Arthropoda' and analysis of Hox genes. Whole transcriptome annotation was conducted following the Trinotate pipeline, by predicting proteins with TransDecoder v. 5.2.0 (Bryant et al., 2017) and searching best hits with blastx and blastp against Swiss-Prot and Pfam databases (Consortium, 2018).

Gene identification

Hox genes (*labial*, *proboscipedia*, *Hox3*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, *Antennapedia*, *Ultrabithorax*, *abdominal-A* and *Abdominal-B*), *Distal-less*, *dachshund*, and *homothorax* orthologs were identified from the embryonic transcriptome of *Phrynus marginemaculatus* using BLAST searches (tblastn) with peptide queries of the homologs of *Drosophila melanogaster* and *Parasteatoda tepidariorum*. The best hits were reciprocally blasted to NCBI database and further assessed with a phylogenetic analysis.

For the phylogenetic analysis of Hox genes, we selected arthropod terminals with available genomes and well annotated references: the fruit fly *Drosophila melanogaster* (FlyBase), the beetle *Tribolium castaneum* (GCA_000002335.3) (Insecta), the centipede *Strigamia maritima*

(GCA_000239455.1) (Myriapoda), the tick *Ixodes scapularis* (GCA_002892825.2), the mite *Tetranychus urticae* (GCA_000239435.1), the spider *Parastetoda tepidarioum* (GCA_000365465.2) and the scorpion *Centruroides sculpturatus* (GCA_000671375.2) (Chelicerata) (Chipman et al., 2014; Grbić et al., 2011; Gulia-Nuss et al., 2016; Herndon et al., 2020; Schwager et al., 2017) (Additional file 5: Table S1). For the analysis of *Distal-less*, *dachshund*, and *homothorax* we selected terminals used in Nolan et al. (Nolan et al., 2020) and included the peptide sequence for *P. marginemaculatus* candidates (Additional file 5: Table S1). Peptides were predicted with ORF Finder (NCBI), TransDecoder v. 5.5.0 (Bryant et al., 2017), or obtained as CDS when possible from publicly available NCBI database. Peptide sequences were aligned using Clustal Omega (Sievers et al., 2011) in the software SeaView v. 4 (Gouy et al., 2010). Gene trees were inferred under maximum likelihood using IQ-TREE v. 1.6.8 (Nguyen et al., 2015) with automatic model selection (-m TEST) and 1000 ultrafast bootstrap resampling (-bb 1000). Alignments for the Hox gene analysis are available in Additional files 6, 7. Genes referred throughout text refer to the longest isoform assembled by Trinity in a contig, and annotated by the procedure detailed above. Paralogs were discerned from gene fragments using as criteria the presence of largely overlapping sequences with amino acid differences.

Fixation and in situ hybridization

Collecting freshly laid eggs and keeping them in controlled temperature with the female or incubator (26 °C) allows to select desirable stages for fixation. We fixed embryos for in situ hybridization with three methods: (1) intact eggs in a phase of 4% formaldehyde/PBS and heptane overnight, after an established spider fixation protocol (Akiyama-Oda and Oda, 2003); (2) piercing a hole in the vitelline membrane, and immersing eggs in 4% formaldehyde/PBS overnight; (3) dissecting the embryo out of the vitelline membrane and 2 to 24 h. All three

procedures were conducted in a shaking platform at room temperature (~ 22 °C). Following the fixation period, embryos were rinsed several times in 1× PBS/0.1% Tween-20 (Sigma Aldrich) (PBS-T), and gradually dehydrated into pure methanol. Embryos were then stored in 20 °C freezer until next procedures.

cDNA was synthesized with SuperScriptIII kit (Thermo Fisher) following the manufacturer's instructions, using the same total RNA used to generate the embryonic transcriptome. Riboprobe templates for the genes *Pmar-dac1*, *Pmar-dac2*, *Pmar-hth1*, and *Pmar-hth-2* in situ hybridizations were generated as follows: genes were PCR amplified from cDNA using gene-specific primers designed with Primer3 v. 4.1.0 (Koressaar and Remm, 2007) and complemented with T7 linker sequences (5′-ggccgcgg-3′; 5′-cccggggc-3′), and a second PCR using universal T7 primer to generate T7 polymerase templates specific for sense (control) and anti-sense (signal) probes (Additional file 8: Table S2). Sense and antisense riboprobes for *Pmar-Dll* were generated with T7 and T3 polymerases using a plasmid template. For cloning of *Pmar-Dll*, gene-specific primers were used to amplify an 848 bp fragment (Additional file 8: Table S2), which was incorporated to vectors and *E. coli* using TOPO TA cloning kit (Thermo Fisher) following the manufacturer's instructions, and Sanger sequenced to confirm their identities.

For colorimetric in situ hybridization, we adapted a protocol for the spider *Parasteatoda tepidariorum* after Akiyama-Oda and Oda (Akiyama-Oda and Oda, 2003). The staining reactions lasted between 2 and 8 h at room temperature using nitro-blue tetrazolium (NBT) and 5-bromo-4-chloro- 3′-indolyphosphate (BCIP) staining reactions (Sigma Aldrich). Embryos were then rinsed in PBS-T (0.05% Tween-20), counter-stained with 10 μg/mL Hoechst 33342 (Sigma-Aldrich, St. Louis, MO, USA), post-fixed in 4% formaldehyde/PBS-T and stored at 4 °C. Images were taken in a petri dish with agar under PBS-T, using a Nikon SMZ25 fluorescence stereomicroscope mounted with a DS-Fi2 digital color camera (Nikon Elements

software). Appendage mounts were imaged in 70% glycerol/ PBS-T in a Zeiss LCM 710 confocal microscope, or with an Olympus DP70 color camera mounted on an Olympus BX60 epifluorescence compound microscope.

Supplementary information

Supplementary information accompanies this paper at https://doi. org/10.1186/s13227-020-00163-w.

Additional file 1: Fig. S1. Tree topology inferred from maximum likelihood analysis of a conserved region (75 amino acid characters) (ln L=−4780.167) using the same terminals of Fig. 3 and including *Parasteatoda tepidariorum Hox3-A* paralog and *Drosophila melanogaster bicoid* (*Hox3*) (red arrow). Numbers on the notes are ultrafast bootstrap resampling frequencies (only > 80 shown). For abbreviations, see Additional file 5: Table S1.

Additional file 2: Fig. S2. Tree topology of *Distal-less (Dll)* inferred from maximum likelihood analysis of amino acid sequences (ln L=−15,272.780). Numbers on the notes are ultrafast bootstrap resampling frequencies (only>80 shown). Accession numbers are available in Additional file 5: Table S1. The terminal for the whip spider *Phrynus marginemaculatus Dll* ortholog is in boldface.

Additional file 3: Fig. S3. Tree topology of *dachshund (dac)* inferred from maximum likelihood analysis of amino acid sequences (ln L=−21,699.907). Numbers on the notes are ultrafast bootstrap resampling frequencies (only>80 shown). The terminals for the whip spider *Phrynus marginemaculatus dac* orthologs are in boldface. The original protein alignment of all terminals (except *P. marginemaculatus*) is from Nolan et al. (2020).

Additional file 4: Fig. S4. Tree topology of *homothorax* (*hth*) inferred from maximum likelihood analysis of amino acid sequences (ln L=−10,459.647). Numbers on the notes are ultrafast bootstrap resampling frequencies (only>80 shown). The terminals for the whip spider

Phrynus marginemaculatus hth orthologs are in boldface. The original protein alignment of all terminals (except *P. marginemaculatus*) is from Nolan et al. (2020).

Additional file 5: Table S1. List of species, accession numbers and abbreviations of terminals used in the phylogenetic analysis of Hox genes, *Distal-less, dachshund* and *homothorax* homologs.

Additional file 6: Protein alignment used in the phylogenetic analysis depicted in Fig. 3.

Additional file 7: Protein alignment used in the phylogenetic analysis depicted in Additional file 1: Figure S1.

Additional file 8: Table S2. List of primers for *Pmar-Dll, Pmar-dac1/2*, *Pmar-hth1/2*, and cloned fragments for *Pmar-Dll*.

Acknowledgements

This work is dedicated to Peter Weygoldt, whose pioneer work in the development of *Phrynus marginemaculatus* and contributions to the study of Amblypygi biology continue to inspire the authors of this paper. Comments from two anonymous reviewers improved an early draft of this manuscript. Sequencing was performed at the University of Wisconsin-Madison BioTechnology Center. Microscopy was performed at the Newcomb Imaging Center, Department of Botany, University of Wisconsin-Madison.

Authors' contributions

Both authors designed the study, analyzed data, developed protocols and contributed to writing. GG conducted gene searches, orthology assessment, phylogenetic analysis, in situ hybridization, assembled figures and drafted the manuscript. Both authors read and approved the final manuscript.

Funding

GG was supported by a Wisconsin Alumni Research Foundation Fall Research Competition award. Materials for the study were supported by an Oscar Franke student competition award to GG from the International Society of Arachnology. This material is based on work supported by the National Science Foundation under Grant IOS-1552610 to PPS.

Availability of data and materials

The data generated and analyzed in this study is included in the main article, additional information and available upon request. Raw reads are deposited in NCBI Sequence Read Archive under accession SRR12232018.

Ethics approval and consent to participate

Studies with arachnids do not require ethics approval or consent to participate.

Consent for publication

Not applicable.

Competing interests

The authors declare having no conflict of interest.

Received: 8 July 2020 Accepted: 11 August 2020

Published online: 28 August 2020

References

Abzhanov, A. and Kaufman, T. C. (2000). Homologs of *Drosophila* appendage genes in the patterning of arthropod limbs. *Dev. Biol.* 227, 673–689.

Akiyama-Oda, Y. and Oda, H. (2003). Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* 130, 1735–1747.

Akiyama-Oda, Y. and Oda, H. (2006). Axis specification in the spider embryo: *dpp* is required for radial-to-axial symmetry transformation and *sog* for ventral patterning. *Development* 133, 2347–2357.

Akiyama-Oda, Y. and Oda, H. (2010). Cell migration that orients the dorsoventral axis is coordinated with anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. *Development* 137, 1263–1273.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440–22.

Barnett, A. A. and Thomas, R. H. (2012). The delineation of the fourth walking leg segment is temporally linked to posterior segmentation in the mite *Archegozetes longisetosus* (Acari: Oribatida, Trhypochthoniidae). *Evol. Dev.* 14, 383–392.

Barnett, A. A. and Thomas, R. H. (2013a). Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *Evodevo* 4, 23.

Barnett, A. A. and Thomas, R. H. (2013b). The expression of limb gap genes in the mite *Archegozetes longisetosus* reveals differential patterning mechanisms in chelicerates. *Evol. Dev.* 15, 280–292.

Beck, L., Foelix, R., Gödeke, E. and Kaiser, R. (1977). Morphology, larval development, and hair sensilla of the antenniform legs of the whip spider *Heterophrynus longicornis* Butler (Arach., Amblypygi). *Zoomorphologie* 88, 259–276.

Blackburn, D. C., Conley, K. W., Plachetzki, D. C., Kempler, K., Battelle, B.-A. and Brown, N. L. (2008). Isolation and expression of *Pax6* and *atonal* homologues in the American horseshoe crab, *Limulus polyphemus*. *Dev. Dyn.* 237, 2209–2219.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell. Rep.* 18, 762–776.

Casares, F. and Mann, R. S. (1998). Control of antennal versus leg development in *Drosophila*. *Nature* 392, 723–726.

Chapin, K. J. and Hebets, E. A. (2016). The behavioral ecology of amblypygids. *J. Arachnol.* 44, 1–14.

Chipman, A. D., Ferrier, D. E. K., Brena, C., Qu, J., Hughes, D. S. T., Schröder, R., Torres-Oliva, M., Znassi, N., Jiang, H., Almeida, F. C., et al. (2014). The first myriapod genome sequence reveals conservative arthropod gene content and genome organization in the centipede *Strigamia maritima*. *PLOS Biol.* 12, e1002005–24.

Consortium, T. U. (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.

Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. and Tautz, D. (1998). A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a

spider. *Proc. Natl. Acad. Sci. USA* 95, 10665–10670.
Dearden, P. K. (2002). Expression of pair-rule gene homologues in a chelicerate: early patterning of the two-spotted spider mite *Tetranychus urticae*. *Development* 129, 5461–5472.
Dong, P. D., Chu, J. and Panganiban, G. (2001). Proximodistal domain specification and interactions in developing *Drosophila* appendages. *Development* 128, 2365–2372.
Dong, P. D. S., Dicks, J. S. and Panganiban, G. (2002). *Distal-less* and *homothorax* regulate multiple targets to pattern the *Drosophila* antenna. *Development* 129, 1967–1974.
Duncan, D. M., Burgess, E. A. and Duncan, I. (1998). Control of distal antennal identity and tarsal development in *Drosophila* by *spineless-aristapedia*, a homolog of the mammalian dioxin receptor. *Genes Dev.* 12, 1290–1303.
Foelix, R. F. (1975). Occurrence of synapses in peripheral sensory nerves of arachnids. *Nature* 254, 146–148.
Foelix, R. F., Chu-Wang, I. W. and Beck, L. (1975). Fine structure of tarsal sensory organs in the whip spider *Admetus pumilio* (Amblypygi, Arachnida). *Tissue Cell* 7, 331–346.
Foelix, R., Troyer, D. and Igelmund, P. (2002). Peripheral synapses and giant neurons in whip spiders. *Microsc. Res. Tech.* 58, 272–282.
Fowler-Finn, K. D. and Hebets, E. A. (2006). An Examination of agonistic interactions in the whip spider *Phrynus marginemaculatus* (Arachnida, Amblypygi). *J. Arachnol.* 34, 62–76.
Gainett, G., Ballesteros, J. A., Kanzler, C. R., Zehms, J. T., Zern, J. M., Aharon, S., Gavish-Regev, E. and Sharma, P. P. (2020). How spiders make their eyes: Systemic paralogy and function of retinal determination network homologs in arachnids. *bioRxiv* 129, 1143–34.
Garwood, R. J., Dunlop, J. A., Knecht, B. J. and Hegna, T. A. (2017). The phylogeny of fossil whip spiders. *BMC Evol. Biol.* 17, 1–14.
Giribet, G. (2018). Current views on chelicerate phylogeny—A tribute to Peter Weygoldt. *Zool. Anz. J. Comp. Zool.* 273, 7–13.
Gough, L. H. (1902). The Development of *Admetus Pumilio* Koch: A Contribution to the Embryology of the Pedipalpi. *Q. J. Microsc. Sci. New Ser.* 45, 595–630.
Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.
Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
Grbić, M., Khila, A., Lee, K.-Z., Bjelica, A., Grbić, V., Whistlecraft, J., Verdon, L., Navajas, M. and Nagy, L. (2007). Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism. *BioEssays* 29, 489–496.
Grbić, M., Leeuwen, T. Van, Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.
Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., Sattelle, D. B., Fuente, J. D. La, Ribeiro, J. M., Megy, K., et al. (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* 7, 10507.
Harper, A., Baudouin-Gonzalez, L., Schönauer, A., Seiter, M., Holzem, M., Arif, S., McGregor, A. P. and Sumner-Rooney, L. (2020). Widespread retention of ohnologs in key developmental gene families following whole genome duplication in

arachnopulmonates. *bioRxiv* 160, 1067–40.
Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R. M., Liang, C., Damm, C., Siemanowski, J., et al. (2020). Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genom.* 21, 1–13.
Heymons, R. (1904). Ueber die Entwicklungsgeschichte und Morphologie der Solifugen. *Compte-Rendu des Séances du Sixiéme Congrés Internatinal de Zoologie* 429–436.
Hughes, C. L. and Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* 4, 459–499.
Hughes CL and Kaufman TC (2002). RNAi analysis of *Deformed, proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the Hemipteran head. *Development* 127, 3683–3694.
Igelmund, P. (1987). Morphology, sense organs, and regeneration of the forelegs (whips) of the whip spider *Heterophrynus elaphus* (Arachnida, Amblypygi). *J. Morphol.* 193, 75–89.
Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., et al. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116, 190–199.
Khadjeh, S., Turetzek, N., Pechmann, M., Schwager, E. E., Wimmer, E. A., Damen, W. G. M. and Prpic, N. M. (2012). Divergent role of the Hox gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl Acad. Sci. USA* 109, 4921–4926.
Khila, A. and Grbić, M. (2007). Gene silencing in the spider mite *Tetranychus urticae*: dsRNA and siRNA parental silencing of the *Distal-less* gene. *Dev. Genes Evol.* 217, 241–251.
Koressaar, T. and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.
Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N.-M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018). Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* 35, 2240–2253.
Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D. and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* 10, 2295.
Miranda, G. S. de, Giupponi, A. P. L., Prendini, L. and Scharff, N. (2018). *Weygoldtia*, a new genus of Charinidae Quintero, 1986 (Arachnida, Amblypygi) with a reappraisal of the genera in the family. *Zool. Anz. J. Comp. Zool.* 273, 23–32.
Mittmann, B. and Scholtz, G. (2001). *Distal-less* expression in embryos of *Limulus polyphemus* (Chelicerata, Xiphosura) and *Lepisma saccharina* (Insecta, Zygentoma) suggests a role in the development of mechanoreceptors, chemoreceptors, and the CNS. *Dev. Genes Evol.* 211, 232–243.
Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
Nolan, E. D., López, C. E. S. and Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Dev. Genes Evol.* 230, 137–153.
Nong, W., Qu, Z., Li, Y., Barton-Owen, T., Wong, A. Y. P., Yip, H. Y., Lee, H. T., Narayana, S., Baril, T., Swale, T., et al. (2020). Horseshoe crab genomes reveal the evolutionary fates of genes and microRNAs after three rounds (3R) of whole genome

duplication. *bioRxiv* 6, 2020.04.16.045815.
Nossa, C. W., Havlak, P., Yue, J.-X., Lv, J., Vincent, K. Y., Brockmann, H. J. and Putnam, N. H. (2014). Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* 3, 708–21.
Oda, H. and Akiyama-Oda, Y. (2008). Differing strategies for forming the arthropod body plan: lessons from Dpp, Sog and Delta in the fly *Drosophila* and spider *Achaearanea*. *Dev. Growth Differ.* 50, 203–214.
Pace, R. M., Grbić, M. and Nagy, L. M. (2016). Composition and genomic organization of arthropod Hox clusters. *Evodevo* 7, 1–11.
Paese, C. L. B., Schoenauer, A., Leite, D. J., Russell, S. and McGregor, A. P. (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *Elife* 7, 1735.
Pechmann, M. and Prpic, N.-M. (2009). Appendage patterning in the South American bird spider *Acanthoscurria geniculata* (Araneae: Mygalomorphae). *Dev. Genes Evol.* 219, 189–198.
Pechmann, M., Schwager, E. E., Turetzek, N. and Prpic, N.-M. (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc. R. Soc. B* 282, 20151162.
Pechmann, M., Benton, M. A., Kenny, N. J., Posnien, N. and Roth, S. (2017). A novel role for *Ets4* in axis specification and cell migration in the spider *Parasteatoda tepidariorum*. *Elife* 6, e27590.
Pereyaslawzewa, S. (1901). Développement embryonnaire des phrynes. *Ann. Sci. Nat. Zool.* 13, 104–117.
Prpic, N.-M. and Damen, W. M. (2004). Expression patterns of leg genes in the mouthparts of the spider *Cupiennius salei* (Chelicerata: Arachnida). *Dev. Genes Evol.* 214, 1–7.
Prpic, N.-M., Janssen, R., Wigand, B., Klingler, M. and Damen, W. G. M. (2003). Gene expression in spider appendages reveals reversal of *exd/hth* spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. *Dev. Biol.* 264, 119–140.
Rayor, L. S. and Taylor, L. A. (2006). Social behavior in amblypygids, and a reassessment of arachnid social patterns. *J. Arachnol.* 34, 399–421.
Ronshaugen, M., McGinnis, N. and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914–917.
Samadi, L., Schmid, A. and Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *Evodevo* 6, 16.
Santer, R. D. and Hebets, E. A. (2008). Agonistic signals received by an arthropod filiform hair allude to the prevalence of near-field sound communication. *Proc. R. Soc. B* 275, 363–368.
Santer, R. D. and Hebets, E. A. (2009). Tactile learning by a whip spider, *Phrynus marginemaculatus* C.L. Koch (Arachnida, Amblypygi). *J. Comp. Physiol. A.* 195, 393–399.
Santer, R. D. and Hebets, E. A. (2011a). Evidence for Air Movement Signals in the Agonistic Behaviour of a Nocturnal Arachnid (Order Amblypygi). *PLoS One* 6, e22473.
Santer, R. D. and Hebets, E. A. (2011). The Sensory and Behavioural Biology of Whip Spiders (Arachnida, Amblypygi). In *Advances in Insect Physiology* (ed. Casas, J.), pp. 1–64. Burlington: Academic Press.
Santos, V. T., Ribeiro, L., Fraga, A., Barros, C. M., Campos, E., Moraes, J., Fontenele, M. R., Araújo, H. M., Feitosa, N. M., Logullo, C., et al. (2013). The embryogenesis of the tick *Rhipicephalus* (*Boophilus*) *microplus*: the establishment of a new chelicerate

model system. *Genesis* 51, 803–818.
Schomburg, C., Turetzek, N. and Prpic, N.-M. (2020). Candidate gene screen for potential interaction partners and regulatory targets of the Hox gene labial in the spider *Parasteatoda tepidariorum*. *Dev. Genes Evol.* 230, 1–16.
Schoppmeier, M. and Damen, W. G. M. (2001). Double-stranded RNA interference in the spider *Cupiennius salei*: the role of *Distal-less* is evolutionarily conserved in arthropod appendage formation. *Dev. Genes Evol.* 211, 76–82.
Schwager, E. E., Schoppmeier, M., Pechmann, M. and Damen, W. G. M. (2007). Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* 4, 10–11.
Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P. and Damen, W. G. M. (2009). *hunchback* functions as a segmentation gene in the spider *Achaearanea tepidariorum*. *Curr. Biol.* 19, 1333–1340.
Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.
Seiter, M., Lemell, P., Gredler, R. and Wolff, J. O. (2019). Strike kinematics in the whip spider *Charon* sp. (Amblypygi: Charontidae). *J. Arachnol.* 47, 260–7.
Setton, E. V. W. and Sharma, P. P. (2018). Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proc. Natl. Acad. Sci. USA* 128, 201720193–10.
Setton, E. V. W., March, L. E., Nolan, E. D., Jones, T. E., Cho, H., Wheeler, W. C., Extavour, C. G. and Sharma, P. P. (2017). Expression and function of *spineless* orthologs correlate with distal deutocerebral appendage morphology across Arthropoda. *Dev. Biol.* 430, 224–236.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012a). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* 14, 522–533.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012b). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.
Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol. Dev.* 15, 228–242.
Sharma, P. P., Schwager, E. E., Extavour, C. G. and Wheeler, W. C. (2014a). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661.
Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C. and Giribet, G. (2014b). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31, 2963–2984.
Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C. and Extavour, C. G. (2015a). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc R. Soc. B* 282, 20150698.
Sharma, P. P., Santiago, M. A., González-Santillán, E., Monod, L. and Wheeler, W. C. (2015b). Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol. Dev.* 17, 347–355.
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E. and Venkatesh, B. (2020a). Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.*

11, 2322.
Shingate, P., Ravi, V., Prasad, A., Tay, B.-H. and Venkatesh, B. (2020b). Chromosome-level genome assembly of the coastal horseshoe crab (*Tachypleus gigas*). *Mol. Ecol. Resour.* 20, 1748–1760.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
Stollewerk, A., Schoppmeier, M. and Damen, W. G. M. (2003). Involvement of *Notch* and *Delta* genes in spider segmentation. *Nature* 423, 863–865.
Struhl, G. (1981). A homoeotic mutation transforming leg to antenna in Drosophila. *Nature* 292, 635–638.
Struhl, G. (1982). Genes controlling segmental specification in the *Drosophila* thorax. *Proc. Natl. Acad. Sci. USA* 79, 7380–7384.
Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* 95, 10671–10675.
Thomas, R. H. and Telford, M. J. (1999). Appendage development in embryos of the oribatid mite *Archegozetes longisetosus* (Acari, Oribatei, Trhypochthoniidae). *Acta Zool.* 80, 193–200.
Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J. and Prpic, N.-M. (2015). Neofunctionalization of a duplicate *dachshund* gene underlies the evolution of a novel leg segment in arachnids. *Mol. Biol. Evol.* 33, 109–121.
Turetzek, N., Khadjeh, S., Schomburg, C. and Prpic, N.-M. (2017). Rapid diversification of *homothorax* expression patterns after gene duplication in spiders. *BMC Evol. Biol.* 17, 168.
Wagner, G. P., Amemiya, C. and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. USA* 100, 14603–14606.
Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V and Zdobnov, E. M. (2017). BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548.
Weygoldt, P. (1969). Beobachtungen zur fortpflanzungsbiologie und zum verhalten der geißelspinne *Tarantula marginemaculata* C. L. Koch (Chelicerata, Amblypygi). *Z. Morph. Tiere* 64, 338–360.
Weygoldt, P. (1970). Lebenszyklus und postembryonale entwicklung der geibelspinne *Tarantula marginemaculata* C. L. Koch (Chelicerata, Amblypygi) im laboratorium. *Z. Morph. Tiere* 67, 58–85.
Weygoldt, P. (1975). Untersuchungen zur Embryologie und Morphologie der Geißelspinne *Tarantula marginemaculata* C. L. Koch (Arachnida, Amblypygi, Tarantulidae). *Zoomorphologie* 82, 137–199.
Weygoldt, P. (2000). *Whip Spiders (Chelicerata: Amblypygi). Their Biology, Morphology and Systematics.* Apollo Books.
Wiegmann, D. D., Hebets, E. A., Gronenberg, W., Graving, J. M. and Bingman, V. P. (2016). Amblypygids: model organisms for the study of arthropod navigation mechanisms in complex environments? *Front. Behav. Neurosci.* 10, 1–8.
Zhou, Y., Liang, Y., Yan, Q., Zhang, L., Chen, D., Ruan, L., Kong, Y., Shi, H., Chen, M. and Chen, J. (2020). The draft genome of horseshoe crab *Tachypleus tridentatus* reveals its evolutionary scenario and well-developed innate immunity. *BMC Genom.* 21, 137–15.

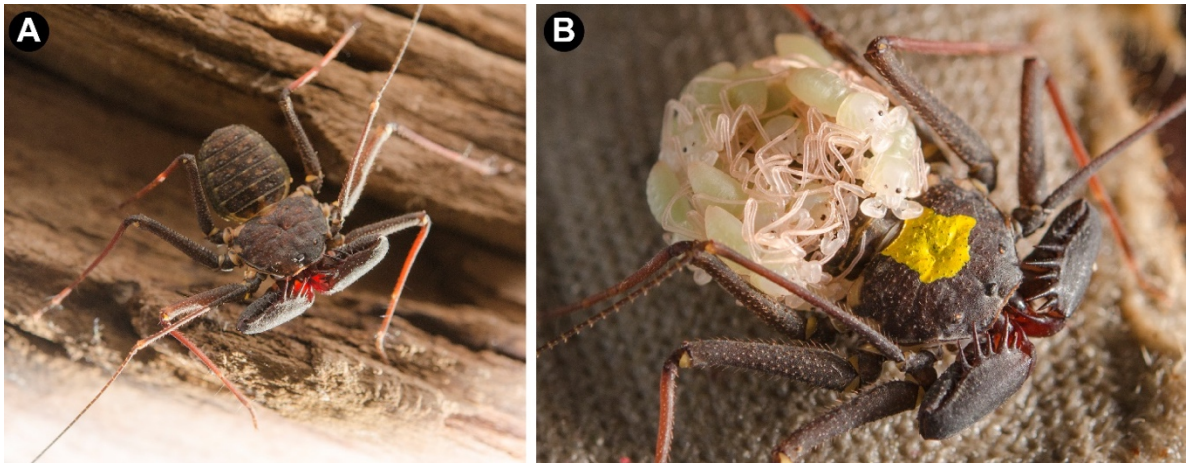


Figure 1: *Phrynus marginemaculatus* as a model to study Amblypygi development. (a) Adult male *P. marginemaculatus*, dorsal view. (b) Adult female *P. marginemaculatus*, carrying praenymphae on the opisthosoma. The yellow ink is used to mark the females in the colony.

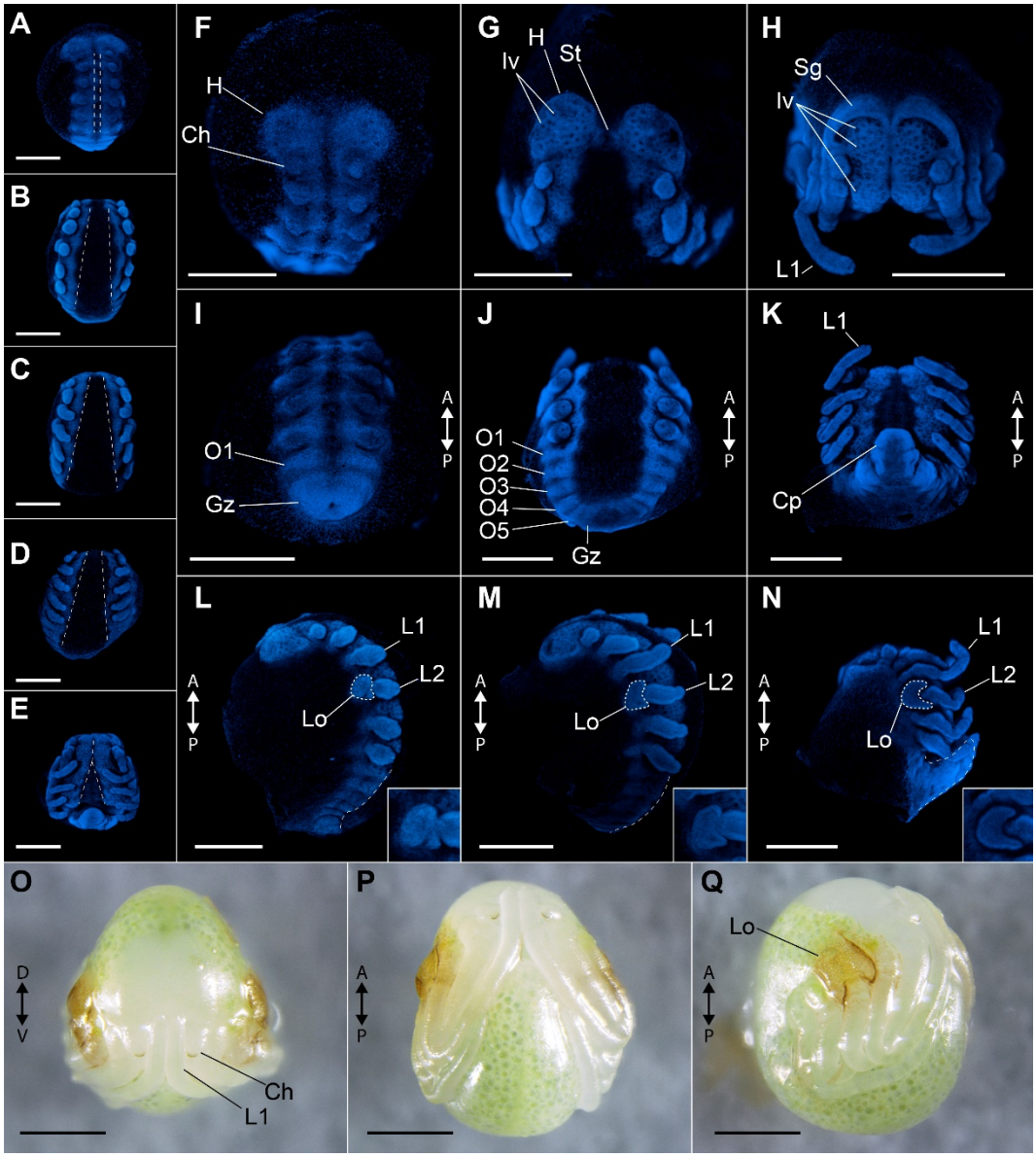


Figure 2. Overview of the embryonic development of *Phrynus marginemaculatus*. Maximum intensity projections of Z-stacks of embryos imaged with nuclear staining (Hoechst) fluorescence. (a–e) Five embryonic stages, in ventral view, spanning the elongation of appendages, and showing germ band inversion and development of the ventral midline. (f–h) Head development, frontal view. (f) Undifferentiated head lobes. (g) Head lobe with neuron precursor cell invaginations. (h) Head lobes with anterior rim folding (semi-lunar grooves). (i–k) Opisthosomal development, posterior view. (i) Embryo with one opisthosomal segment. (j) Embryo with five opisthosomal segments. (k) Embryo with ventrally curved opisthosoma and most opisthosomal segments formed (number unclear). (l), (n) Lateral organ formation and opisthosomal curving, lateral view. (l) Lateral organ as an undifferentiated bud proximal in leg II. (m) Lateral organ with a discreet "C" shape, and opisthosomal segments curved up. n Lateral organ in its final "kidney" shape, and opisthosoma curved ventrally. Insets in (l–n) show the lateral organ (Lo), which is outlined with dashed lines in the respective main panel. Dashed lines on the posterior regions of the embryos in l–n mark the curvature of the opisthosoma. (o–q) Deutembryo, in frontal, ventral and lateral view, respectively. *Ch* chelicera, *Gz* growth zone, *H* head lobe, *Iv* neural precursor cell invagination, *St* stomodeum. Scale bars: 500 µm.

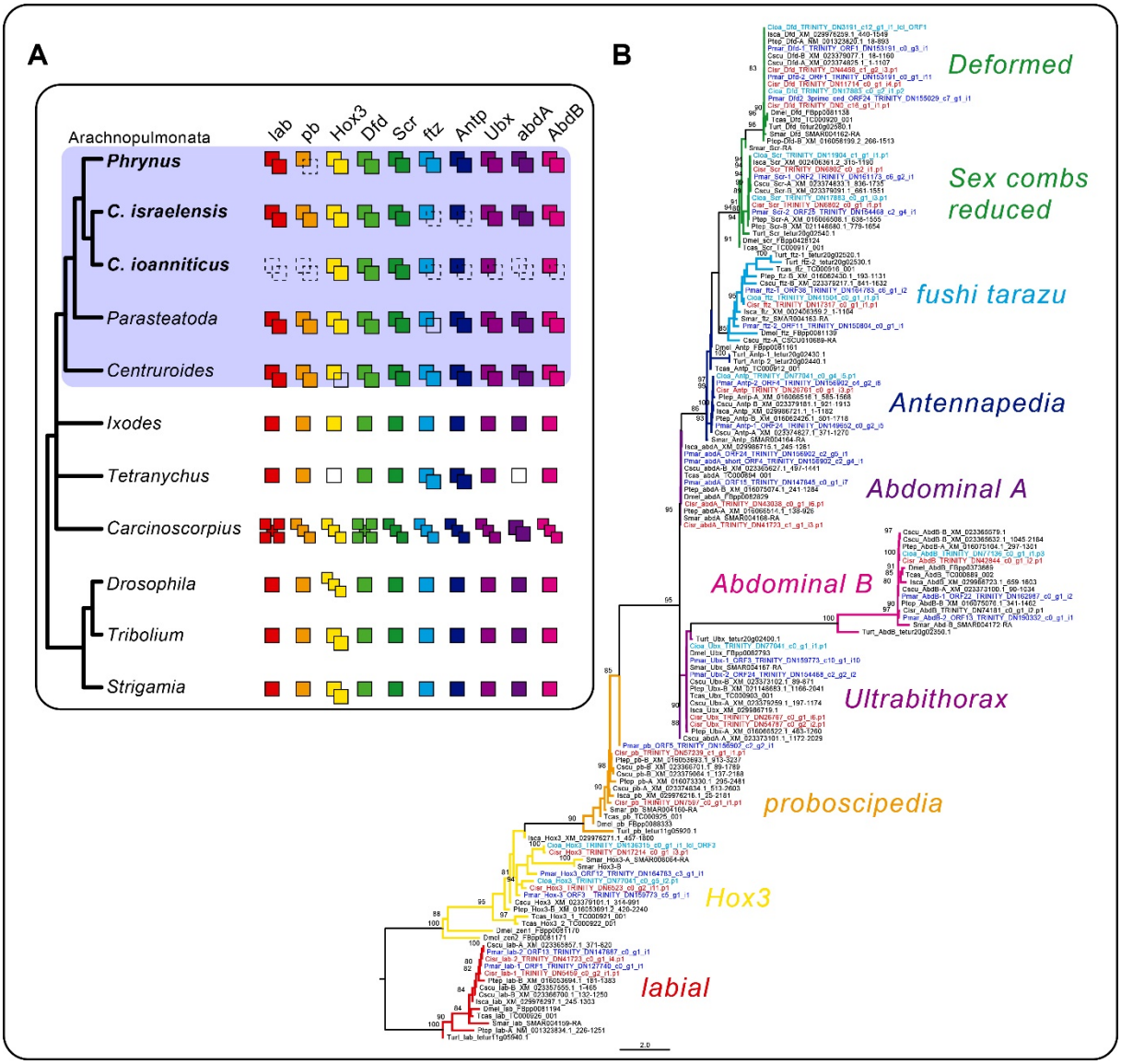


Figure 3. Hox gene duplications in Amblypygi. (a) Schematic representation of the number of copies of each of the ten Hox genes in the selected arthropod terminals. Dotted squares in the amblypygids *Phrynus marginemaculatus*, *Charinus israelensis*, and *C. ioanniticus* indicate absence from the transcriptome. Empty squares in Parasteatoda, Centruroides, and Tetranychus indicate absence from the genome. Number of Hox genes in *Carcinoscorpius rotundicauda* after (Shingate et al., 2020a) (but see [Nong et al., 2020] for evidence of additional copies). (b) Tree topology inferred from maximum likelihood analysis of a conserved region (71 amino acid characters) using the same terminals of the schematics (ln L=−3442.810). Numbers on the notes are ultrafast bootstrap resampling frequencies (only>80% shown). Species: *Phrynus marginemaculatus (Pmar); Charinus israelensis* (*Cisr*)*; C. ioanniticus* (*Cioa*)*; Parasteatoda tepidariorum* (*Ptep*)*; Centruroides sculpturatus* (*Cscu*)*; Ixodes scapularis* (*Isca*)*; Tetranychus urticae* (*Ttur*)*; Strigamia maritima* (*Smar*)*; Drosophila melanogaster* (*Dmel*)*; Tribolium castaneum* (*Tcas*). *lab: labial; pb: proboscipedia/maxillopedia; Hox3*: *Hox3/zerknullt/z2; Dfd: Deformed; Scr: Sex combs reduced; ftz: fushi tarazu*; *Antp: Antennapedia/prothorax-less; Ubx: Ultrabithorax; abdA: abdominal-A; AbdB: Abdominal-B.*

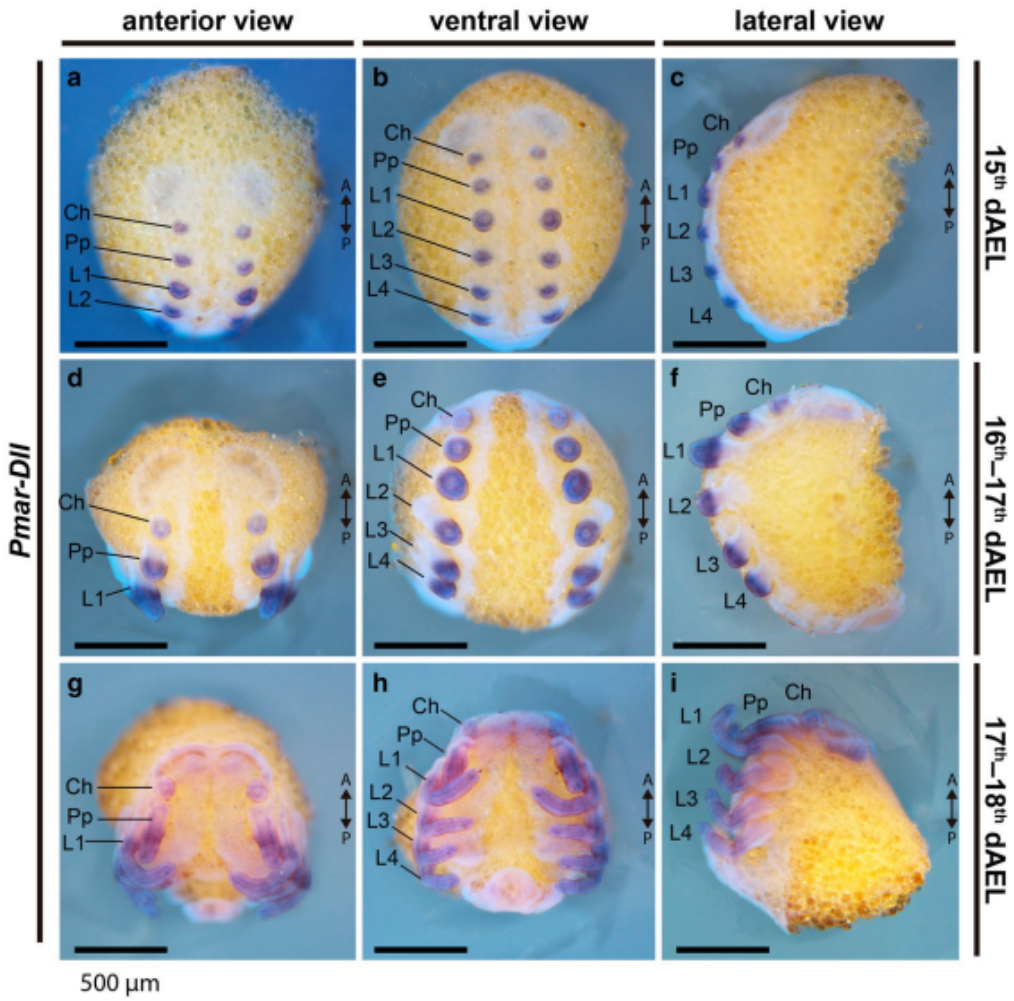


Figure 4. *Pmar-Distal-less* (*Pmar-Dll*) colorimetric in situ hybridization. Whole mount bright field images (Z-stack maximum intensity projection) overlaid with nuclear staining (Hoechst). *Ch* chelicera, *Pp* pedipalp, *L1–4* leg 1–4; *A/P* anterior/posterior, *dAEL* days after egg laying. Scale bars: 500 µm.

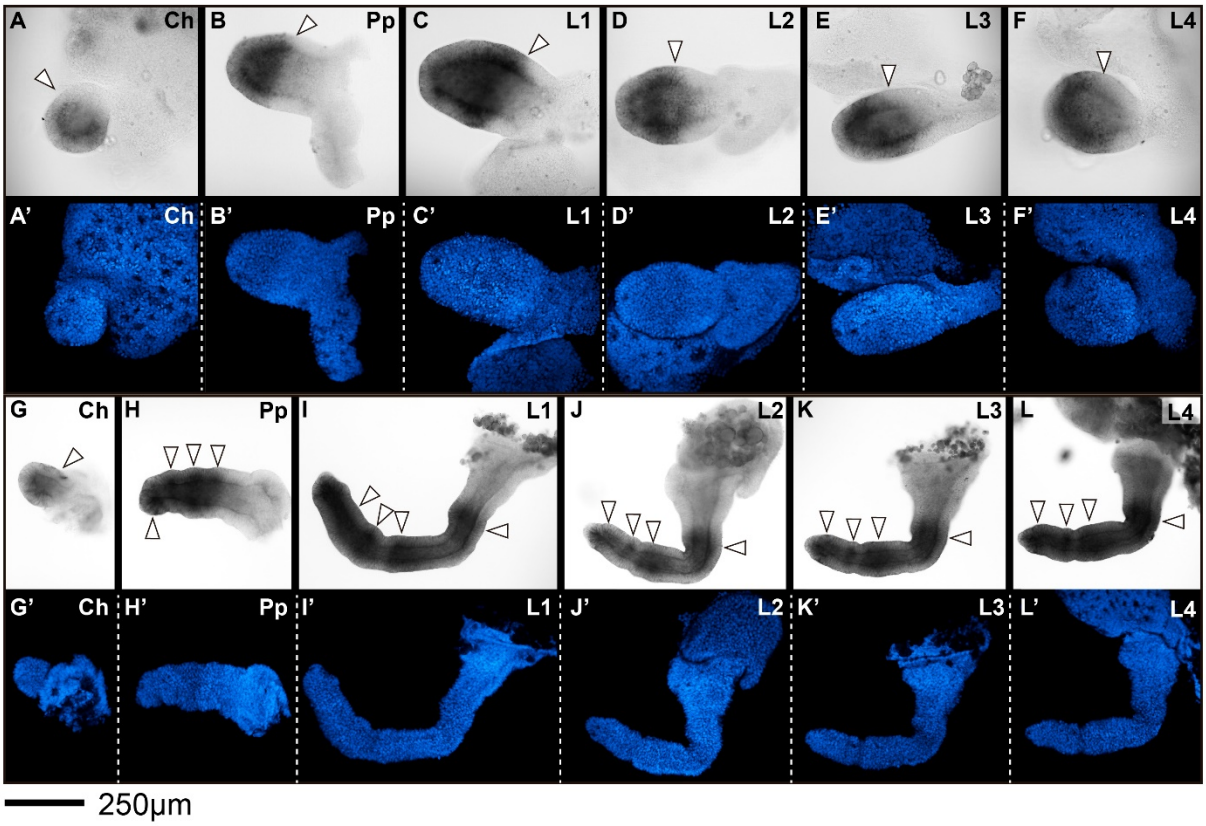


Figure 5. *Pmar-Distal-less* (*Pmar-Dll*) colorimetric in situ hybridization, appendage dissections in lateral view. Distal is to the left. (a–l) Bright field. (a′–l′) Nuclear staining (Hoechst), confocal Z-stack maximum projection. (a–f) 16–17 dAEL embryo. (g–l) 17–18 dAEL embryo. (j), (j)′ have been vertically reflected to adjust orientation. Distal is to the left. Arrowheads in (a–f) indicate proximal boundary of expression. Arrowheads in (g–l) indicate stronger domains of expression. *Ch* chelicera, *Pp* pedipalp, *L1–4* leg 1–4. Scale bar: 250 µm.

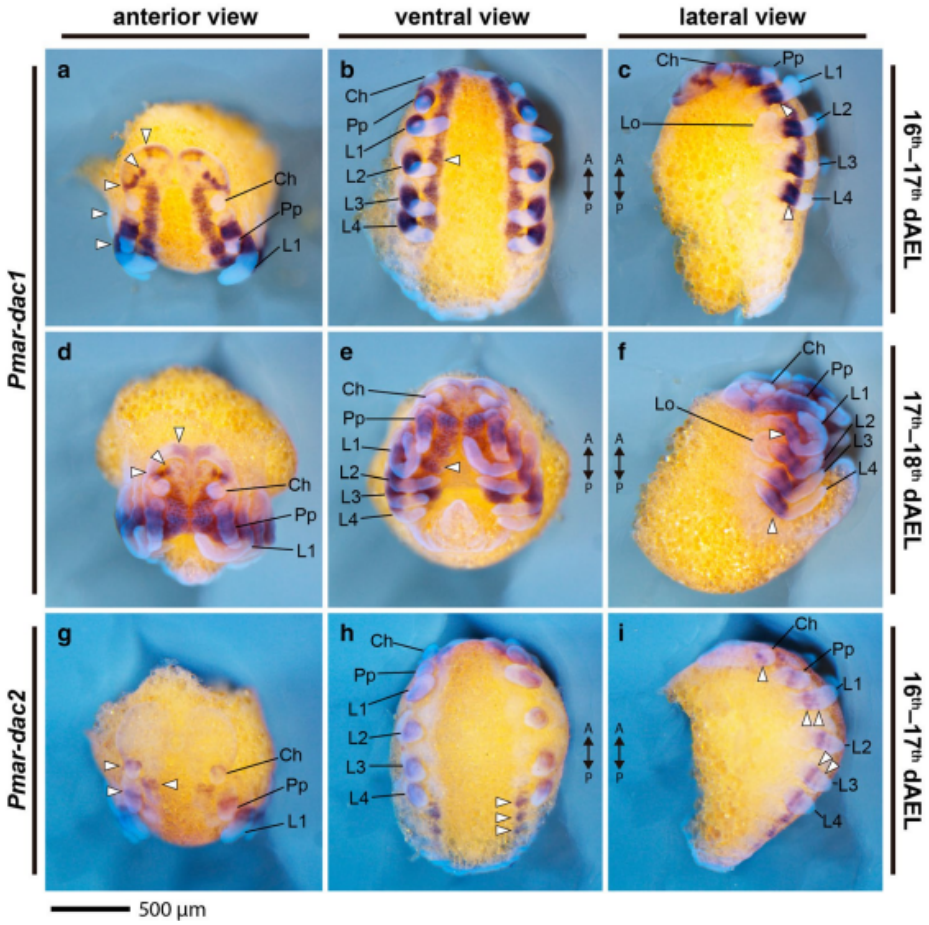


Figure 6. *Pmar-dachshund 1* (*Pmar-dac1*) (a–f) and *Pmar-dachshund 2* (*Pmar-dac2*) (g–i) colorimetric in situ hybridization. Whole mount bright field images (Z-stack automontage) overlaid with nuclear staining (Hoechst). *Ch* chelicera, *Pp* pedipalp, *L1–4* leg 1–4, *Lo* lateral organ, *A/P* anterior/ posterior; arrowhead: expression domain; *dAEL* days after egg laying. Scale bars: 500 µm.

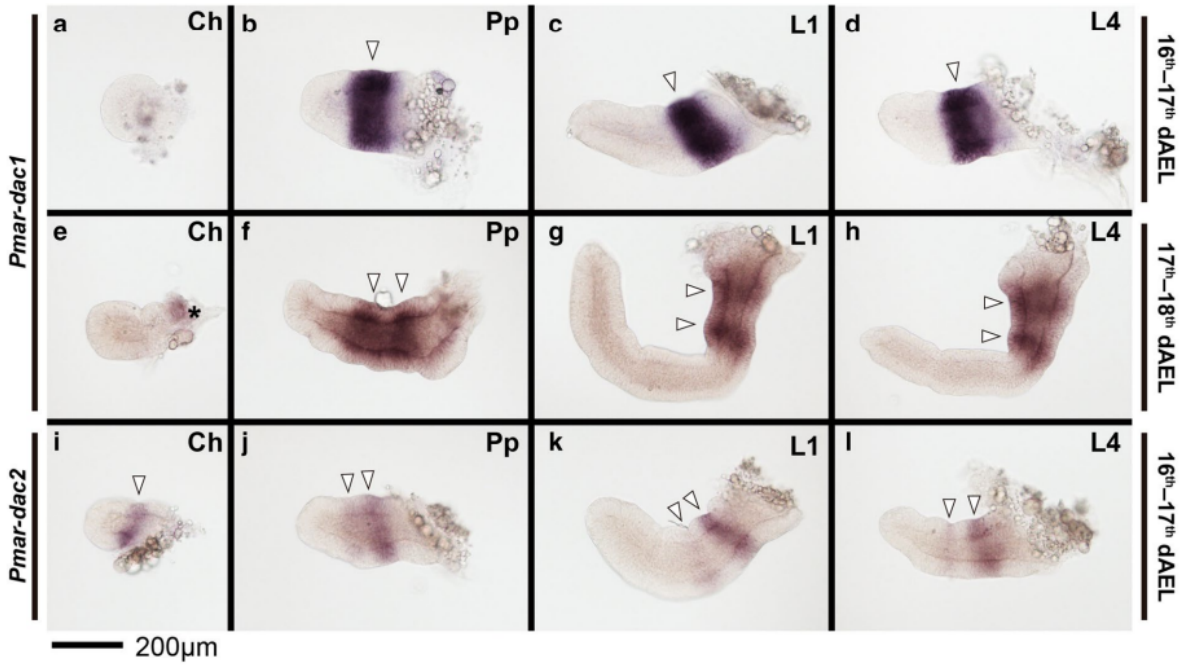


Figure 7. *Pmar-dachshund 1* (*Pmar-dac1*) (a–h) and *Pmar-dachshund 2* (*Pmar-dac2*) (i–l) colorimetric in situ hybridization, appendage dissections in lateral view. Distal is to the left. White arrowhead indicates discreet domains of expression. Asterisk in e marks expression from the lateral margin of the headlobe. *Ch*: chelicera; *Pp*: pedipalp; *L1, 4*: leg 1, 4. *dAEL*: days after egg laying. Scale bar: 200 µm.

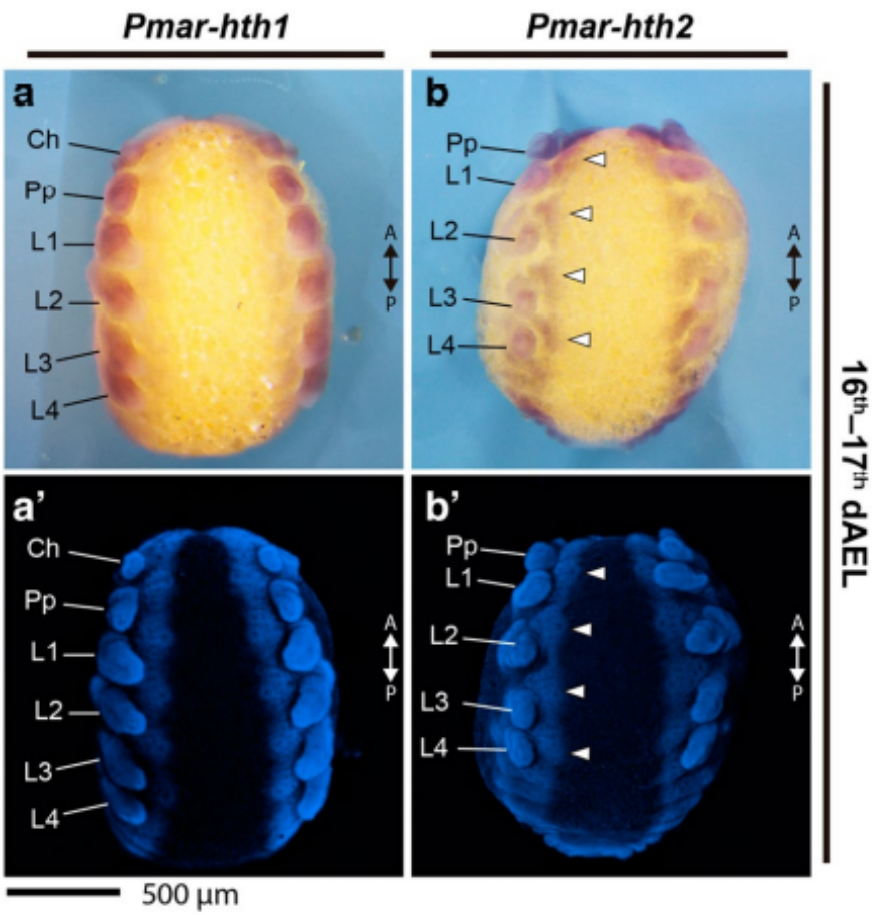


Figure 8. *Pmar-homothorax 1* (*Pmar-hth1*) (a–a′) and *Pmar-homothorax 2* (*Pmar-hth2*) (b–b′) colorimetric in situ hybridization, ventral view. (a), (b) Whole mount bright field images (Z-stack automontage). (a′), (b′) Nuclear staining (Hoechst) (Z-stack automontage). *Ch* chelicera, *Pp* pedipalp, *L1–4* leg 1–4, *A/P* anterior/posterior; arrowhead: expression domain; *dAEL* days after egg laying. Scale bar: 500 µm.

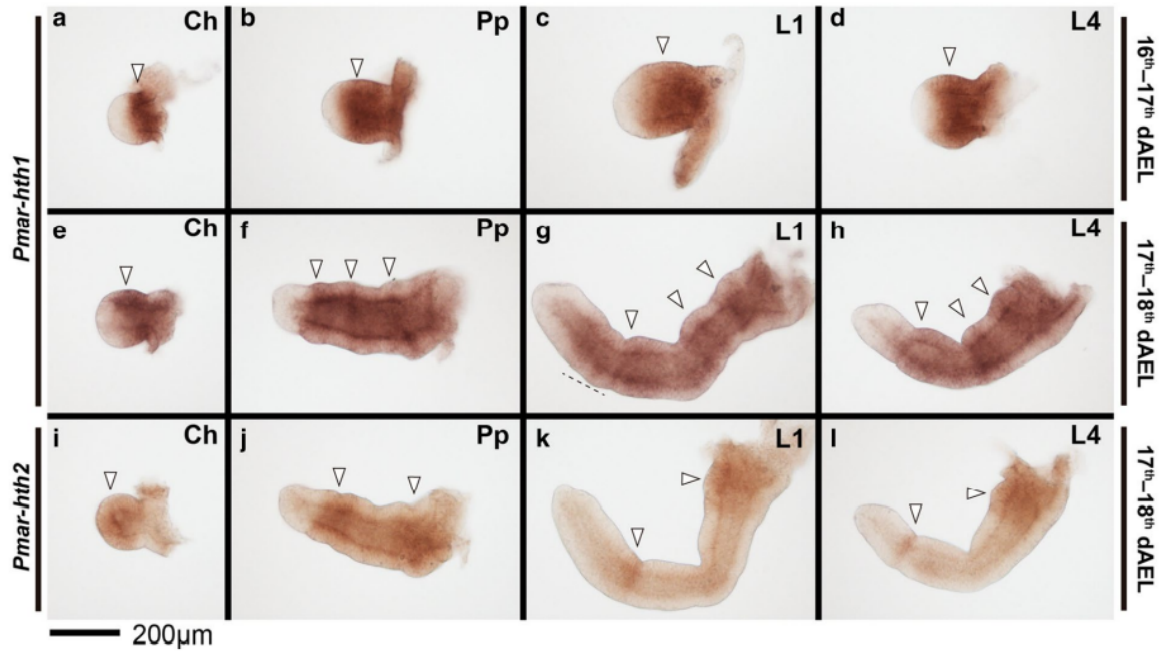


Figure 9. *Pmar-homothorax 1* (*Pmar-hth1*) (a–h) and *Pmar-homothorax 2* (*Pmar-hth2*) (i–l) colorimetric in situ hybridization, appendage dissections in lateral view. Distal is to the left. White arrowhead indicates discreet domains of expression. Dotted line in "G" marks expanded distal expression domain. *Ch*: chelicera; *Pp*: pedipalp; *L1, 4:* leg 1, 4. *dAEL:* days after egg laying. Scale bar: 200 µm.

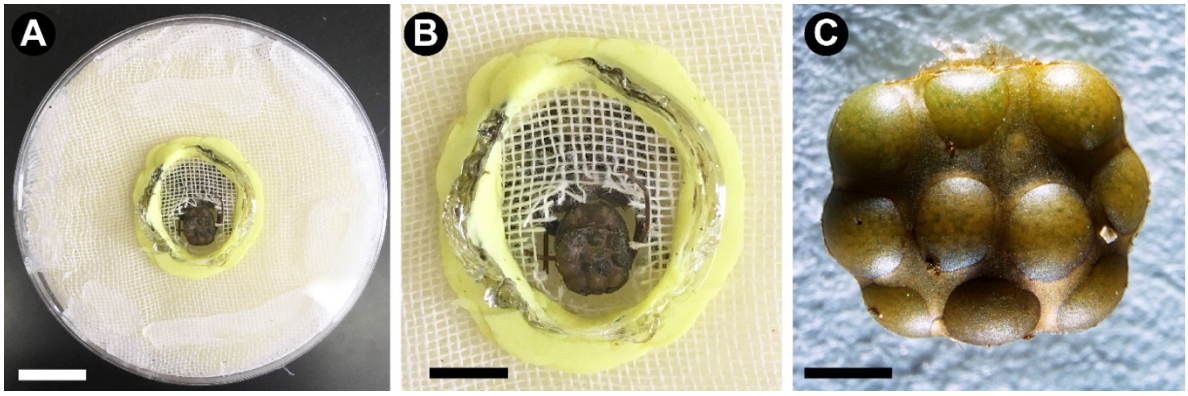


Figure 10. Tool for removing the brood sac from whip spiders. (a) Dissecting dish made of plastic petri dish and foam. (b) Detail of an immobilized female ready to have the brood sac dissected. (c) Intact brood sac dissected from the female. Scale bars: (a) 1 cm (approximate); (b) 5 mm (approximate); (c) 1 mm.

Chapter 5

Systemic paralogy and function of retinal determination network homologs in arachnids

Guilherme Gainett[1†], Jesús A. Ballesteros[1*†], Charlotte R. Kanzler[1], Jakob T. Zehms[1], John M. Zern[1], Shlomi Aharon[2], Efrat Gavish-Regev[2] and Prashant P. Sharma[1]

* Correspondence: ggainett@gmail.com; ballesterosc@wisc.edu
†Guilherme Gainett and Jesús A. Ballesteros contributed equally to this work.
[1]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA
[1]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI 53706, USA.
[2]National Natural History Collections, The Hebrew University of Jerusalem, Jerusalem, 9190401, Israel.

Published November 2020 in BMC Genomics

Volume 21, Article number: 811, doi.org/10.1186/s12864-020-07149-x

Abstract

Background: Arachnids are important components of cave ecosystems and display many examples of troglomorphisms, such as blindness, depigmentation, and elongate appendages. Little is known about how the eyes of arachnids are specified genetically, let alone the mechanisms for eye reduction and loss in troglomorphic arachnids. Additionally, duplication of Retinal Determination Gene Network (RDGN) homologs in spiders has convoluted functional inferences extrapolated from single-copy homologs in pancrustacean models. **Results:** We investigated a sister species pair of Israeli cave whip spiders, *Charinus ioanniticus* and *C. israelensis* (Arachnopulmonata, Amblypygi), of which one species has reduced eyes. We generated embryonic transcriptomes for both Amblypygi species, and discovered that several RDGN homologs exhibit duplications. We show that duplication of RDGN homologs is systemic across arachnopulmonates (arachnid orders that bear book lungs), rather than being a spider-specific phenomenon. A differential gene expression (DGE) analysis comparing the

expression of RDGN genes in field-collected embryos of both species identified candidate RDGN genes involved in the formation and reduction of eyes in whip spiders. To ground bioinformatic inference of expression patterns with functional experiments, we interrogated the function of three candidate RDGN genes identified from DGE using RNAi in the spider *Parasteatoda tepidariorum*. We provide functional evidence that one of these paralogs, sine *oculis/Six1 A* (*soA*), is necessary for the development of all arachnid eye types. **Conclusions:** Our work establishes a foundation to investigate the genetics of troglomorphic adaptations in cave arachnids, and links differential gene expression to an arthropod eye phenotype for the first time outside of Pancrustacea. Our results support the conservation of at least one RDGN component across Arthropoda and provide a framework for identifying the role of gene duplications in generating arachnid eye diversity. **Keywords:** Cave blindness, *sine oculis, Six1*, *Parasteatoda tepidariorum*, Amblypygi, RNAi

Introduction

Cave habitats offer apt systems for investigating the genetic basis of morphological convergence because communities of these habitats are similarly shaped by environmental pressures, such as absence of light and diminished primary productivity [(Howarth, 1993; Juan et al., 2010). Troglobites, species exclusive to cave environments and adapted to life in the dark, exhibit a suite of characteristics common to cave systems around the world, such as reduction or complete loss of eyes, depigmentation, elongation of appendages and sensory structures, and decreased metabolic activity (Jemec et al., 2017; Protas and Jeffery, 2012; Riddle et al., 2018). Previous work has shown that troglomorphism can evolve over short time spans (< 50 kyr) despite gene flow (Bradic et al., 2013; Coghill et al., 2014; Herman et al., 2018) and that parallel evolution of troglomorphic traits (e.g., depigmentation; eye loss) in independent populations can involve the same genetic locus (Protas et al., 2006; Protas et al., 2011; Re et al., 2018).

Troglomorphism and troglobitic fauna have been analyzed across numerous taxonomic groups with respect to systematics and population genetics. However, one component of the troglobitic fauna that remains poorly understood is cave arachnids. Most species of Arachnida are prone to nocturnal habits and some orders broadly exhibit troglophily; in fact, troglobitic species are known from all the extant terrestrial arachnid orders except Solifugae and Thelyphonida (Cruz-López et al., 2016; Esposito et al., 2015; Harvey, 2002; Harvey, 2007; Hedin and Thomas, 2010; López et al., 2014; Mammola et al., 2017; Miranda et al., 2016; Smrž et al., 2013). In addition to eye and pigment loss, troglomorphism in arachnids manifests in the form of compensatory elongation of walking legs and palps, appendages which harbor sensory structures in this group (Derkarabetian et al., 2010; Mammola and Isaia, 2017; Mammola et al., 2018a; Mammola et al., 2018b).

Thorough understanding of the developmental genetic basis for the evolution of

troglomorphic traits has been largely spearheaded in two model systems: the Mexican cavefish *Astyanax mexicanus* (Bradic et al., 2013; Coghill et al., 2014; Herman et al., 2018; Porter et al., 2007; Protas and Jeffery, 2012; Protas et al., 2006) and the cave isopod *Asellus aquaticus* (Jemec et al., 2017; Re et al., 2018; Stahl et al., 2015). Both model systems have more than one hypogean population, can be maintained in the laboratory, and are amenable to approaches such as genetic crosses and quantitative trait locus mapping. The advent of short read sequencing technology in tandem with experimental approaches has transformed the potential to triangulate regulatory differences between hypogean (subterranean) and epigean (surface-dwelling) lineages (Protas et al., 2006; Re et al., 2018; Riddle et al., 2018; Stahl et al., 2015), and to study a broader range of cave taxa. Among arthropods, work on the isopod *A. aquaticus* in particular has made significant advances in the identification of loci regulating pigmentation and size of arthropod eyes (Protas et al., 2011; Re et al., 2018), complementing forward and reverse genetic screening approaches in other pancrustacean models (e.g., *Drosophila melanogaster*, *Tribolium castaneum*, and *Gryllus bimaculatus*) (Cagan, 2009; Kumar, 2009; Takagi et al., 2012; ZarinKamar et al., 2011). However, developmental and genetic insights into the evolution of blindness illuminated by *A. aquaticus* and other pancrustacean models are not directly transferable to Arachnida for two reasons. First, the eyes of arachnids are structurally and functionally different from those of pancrustaceans. Typically, the main eyes of adult Pancrustacea (e.g., *A. aquaticus*) are a pair of faceted (or apposition) eyes, which are composed of many subunits of ommatidia. In addition, adult Pancrustacea have small median ocelli (typically three in holometabolous insects), often located medially and at the top of the head.

By contrast, extant arachnids lack ommatidia and typically have multiple pairs of eyes arranged along the frontal carapace. All arachnid eyes are simple-lens eyes or ocelli; each eye has a single cuticular lens, below which are a vitreous body and visual cells. The

retina is composed of the visual cells and pigment cells. These eyes are divided in two types, namely the principal eyes and the secondary eyes (Foelix, 2011; Land, 1985). Principal and secondary eyes differ in the orientation of their retina (Homann, 1971): the principal eyes are of the everted type, with the visual cells lying distally, and lack a reflective layer; the secondary eyes are inverted, with the light-sensitive rhabdomeres pointing away from incoming light (analogous to vertebrate eyes). All secondary eyes possess a reflective layer of crystalline deposits called a tapetum, which is responsible for the "eye shine" of spiders. The principal eyes are the median eyes (ME, also known as anterior median eyes). The secondary eyes comprise the anterior lateral eyes (ALE), posterior lateral eyes (PLE), and median lateral eyes (MLE; also known as posterior median eyes) (Fig. 1a) (Foelix, 2011; Land, 1985) (nomenclature used here follows Schomburg et al. 2015). Certain orders and suborders of arachnids have lost one type of eye altogether, with the homology of eyes clarified by the fossil record and embryology (Foelix, 2011; Garwood et al., 2014; Morehouse et al., 2017).

The second concern in extrapolating developmental processes derived from pancrustaceans is that a subset of Arachnida exhibits an ancient shared genome duplication, resulting in numerous paralogs of developmental patterning genes. Recent phylogenetic and comparative genomic works on Arachnida have shown that Arachnopulmonata (Ballesteros and Sharma, 2019; Ballesteros et al., 2019; Sharma et al., 2014b), the clade of arachnids that bear book lungs (e.g., spiders, scorpions, whip spiders), retain duplicates of many key transcription factors, such as homeobox genes, often in conserved syntenic blocks (Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2014a; Sharma et al., 2015a). Many of the ensuing paralogs exhibit non-overlapping expression patterns and a small number have been shown to have subdivided the ancestral gene function (subfunctionalization) or acquired new functions (neofunctionalization) (Leite et al., 2018; Paese et al., 2018; Turetzek et al., 2015).

While comparatively little is known about the genetics of arachnid eye development, gene expression surveys of insect retinal determination gene network (RDGN) homologs of two spiders (*Cupiennius salei* and *Parasteatoda tepidariorum*) have shown that this phenomenon extends to the formation of spider eyes as well (Samadi et al., 2015; Schomburg et al., 2015).

Different paralog pairs (orthologs of *Pax6*, *Six1*, *Six3*, *eyes absent*, *atonal*, *dachshund* and *orthodenticle*) exhibit non-overlapping expression boundaries in the developing eye fields, resulting in different combinations of transcription factor expression in the eye pairs (Samadi et al., 2015; Schomburg et al., 2015). While these expression patterns offer a potentially elegant solution to the differentiation of spider eye pairs, only a few studies with the spider *P. tepidariorum* have attempted to experimentally test the role of these genes in the formation of arachnid eyes. *Ptep-orthodenticle-1* maternal RNA interference (RNAi) knockdown results in a range of anterior defects, including complete loss of the head, which precluded assessment of a role in the formation of the eyes (Pechmann et al., 2009). *Ptep-dac2* RNAi knockdown results in appendage segment defects, but no eye patterning defects were reported by the authors (Turetzek et al., 2015). More recently, a functional interrogation of both *Ptep-Six3* paralogs, focused on labrum development, reported no discernible morphological phenotype, despite a lower hatching rate than controls and disruption of a down- stream target with a labral expression domain (Schacht et al., 2020). Thus, gene expression patterns of duplicated RDGN paralogs have never been linked to eye-related phenotypic outcomes in any arachnopulmonate model. Similarly, the functions of the single-copy orthologs of RDGN genes in groups like mites (Grbić et al., 2007; Telford and Thomas, 1998), ticks (Santos et al., 2013) and harvestmen (Garwood et al., 2014; Sharma et al., 2013; Sharma et al., 2015b) are entirely unexplored, in one case because an otherwise tractable arachnid species lacks eyes altogether (the mite *Archegozetes longisetosus*) (Barnett and Thomas, 2012; Barnett and Thomas, 2013a;

Barnett and Thomas, 2013b; Telford and Thomas, 1998).

Investigating the evolution of eye loss in arachnids thus has the potential to elucidate simultaneously (1) the morphogenesis of a poorly understood subset of metazoan eyes (Foelix, 2011; Morehouse et al., 2017), (2) developmental mechanisms underlying a convergent trait (i.e., eye loss in caves) in phylogenetically distant arthropod groups (Protas and Jeffery, 2012; Re et al., 2018), (3) shared programs in eye development common to Arthropoda (through comparisons with pancrustacean datasets) (Cagan, 2009; Stahl et al., 2015; Takagi et al., 2012; ZarinKamar et al., 2011), and (4) the role of ancient gene duplicates in establishing the diversity of eyes in arachnopulmonates (Leite et al., 2018; Samadi et al., 2015; Schomburg et al., 2015).

As first steps toward these goals, we developed transcriptomic resources for a sister species pair of cave-dwelling *Charinus* whip spiders, wherein one species exhibits typical eye morphology and the other highly reduced eyes (a troglobitic condition). We applied a differential gene expression (DGE) analysis to these data-sets to investigate whether candidate RDGN genes with known expression patterns in model spider species (*C. salei*, *P. tepidariorum*) exhibit differential expression in non-spider arachnopulmonates, as a function of both eye condition and developmental stage. To link bioinformatic inference of expression patterns with functional outcomes, we interrogated the function of three candidate RDGN genes identified from DGE in a model arachnopulmonate, using RNAi in the spider *P. tepidariorum*, which exhibits the same number and types of eyes as whip spiders. We provide functional evidence that one of these candidates, *sine oculis/Six1*, is necessary for the development of all spider eye types.

Results

RDGN gene duplication in *Charinus* whip spiders

To investigate the possible role of Retinal Determination Gene Network (RDGN) genes in eye reduction in naturally occurring cave arachnids, we first generated transcriptomic resources for an empirical case of closely related, non-spider arachnopulmonate sister species pair that constitutes one epigean and one troglobitic species: the whip spider species *Charinus ioanniticus* Kritscher 1959 and *C. israelensis* (Fig. 1 B–C) (Miranda et al., 2016). Whip spiders, arachnopulmonates of the order Amblypygi, are commonly found in cave habitats ranging from rainforests, savannahs and deserts (Weygoldt, 2000). The recently described troglobitic species *C. israelensis* (reduced-eyes) occurs in close proximity to its congener *C. ioanniticus* (normal eyes) in caves in the Galilee, northern Israel (Miranda et al., 2016). Given that the formation of Levantine cave refuges is considerably recent, *C. israelensis* and *C. ioanniticus* are likely sister species with a short time of divergence, an inference supported by their similar morphology (Miranda et al., 2016). We collected ovigerous females from both species in caves in Israel and extracted RNA from embryos (Additional file 1, Table S1). Staging of the embryos and nomenclature of stages follows the description of whip spider embryology for the species *Phrynus marginemaculatus* (Gainett and Sharma, 2020; Weygoldt, 1975). Our sampling focused on the deutembryo, the stage where most external features of the embryo, such as tagmosis and appendages are fully formed, but not the eyes. In *P. marginemaculatus*, the eyes begin to form around 50 dAEL, but the eye spots become externally visible and pigmented only close to hatching (90 dAEL) (Weygoldt, 1975).

For de novo assembly of the embryonic transcriptomes of *C. ioanniticus* and *C. israelensis,* we extracted RNA from all deutembryo stages collected in the field (see Additional file 1, Table S1 for localities and sample explanations). Assemblies include two deutembryo stages before eyespot formation and one deutembryo stage bearing eyespots for *C. ioanniticus*; and two early deutembryo stages for *C. israelensis* (Additional file 1, Fig. S1). The assemblies

of *C. ioanniticus* and *C. israelensis* exhibited an N50 of 1122 bp and 1045 bp, respectively (Additional file 1, Table S2); universal single copy ortholog benchmarking with BUSCO v3.0 (Waterhouse et al., 2017) indicated completeness of 93.8 and 95.2%, respectively.

Amblypygi is inferred to be nested stably in Arachnopulmonata, the clade of arachnids that bear book lungs (Ballesteros and Sharma, 2019; Giribet, 2018; Lozano-Fernandez et al., 2019; Rota-Stabelli et al., 2011; Sharma et al., 2014b). Recent evidence suggests that the common ancestor of arachnopulmonates has undergone a whole- or partial-genome duplication affecting large gene families, such as homeobox genes (Leite et al., 2018; Schwager et al., 2017; Sharma et al., 2015a). The well-documented phylogenetic position of Amblypygi in Arachnopulmonata predicts that genes in RDGN that are duplicated in spiders, should also be duplicated in *Charinus* whip spiders (as well as other arachnopulmonate orders). To test this hypothesis, we performed phylogenetically-informed orthology searches on the newly assembled embryonic transcriptomes of both *Charinus* species, and conducted phylogenetic analysis with orthologs across selected arthropod species. We discovered that homologs of *atonal* (*ato*)*, Pax6*, *dachshund* (*dac*), *sine oculis* (*so*; *Six1*), *Optix* (*Six3*), and *orthodenticle* (*otd*) are duplicated in *Charinus*, whereas *eyegone* (*eyg*) and *eyes absent* (*eya*) occur as single-copy orthologs (these latter two also occurring single-copy in spiders) (Fig. 2). A detailed description of the orthology inference and annotation is available in the Additional file 1, Supplementary Results and Figures S2–S8. While the two copies of *ato* and *Pax6* are inferred to result from shared duplication with other arthropods (Fig. 2), the occurrence of paralogs of *dac*, *Optix*, *otd* and *so* in *Charinus* whip spiders, as well as a scorpion, suggests that the retention of RDGN ohnologs is systemic in Arachnopulmonata.

RDGN gene expression differences related to eye formation in whip spiders: comparing early and late stages of *C. ioanniticus.*

The expression of paralog pairs of *Pax6*, *so*, *Optix*, *eya*, *ato*, *dac*, and *otd* in the developing eyes of the spiders (Samadi et al., 2015; Schomburg et al., 2015) and the occurrence of the same paralogs in *Charinus* whip spiders, suggest that these genes may also be involved in the formation of eyes in whip spiders. We investigated this idea by comparing the expression levels of these RDGN genes in the stages before eye-spot formation versus a stage after eye-spot formation in the eye-bearing whip spider *C. ioanniticus* (henceforth "Comparison 1"; Fig. 3a, d).

We mapped reads of both treatments to the reference transcriptome of *C. ioanniticus* using the quasi-alignment software Salmon v. 1.1.0 (Patro et al., 2017) and conducted a differential gene expression analysis of Comparison 1 using DESeq2 v 1.24.0 (Love et al., 2014) (Additional file 1, Fig. S9). These comparisons showed that *Cioa-otdA*, *Ciao-eya* and *Cioa-soA* are significantly more highly expressed ($p_{adj} < 0.05$) in the stage before eyespot formation in comparison with the stage with eyespots (Fig. 3a). The higher relative expression of both *so* and *eya* in that stage accords with the fact that in the fruit fly *D. melanogaster* they form a protein complex that regulates gene expression in synergy (Pignoni et al., 1997). These relative expression dynamics in whip spiders are also consistent with the overlapping expression patterns of *eya* and *so* paralogs in the eyes of spiders (Samadi et al., 2015; Schomburg et al., 2015). These results highlight the three RDGN genes as promising candidates involved in the formation of eyes in whip spiders.

RDGN gene expression differences related to eye reduction in whip spiders: comparing *C. ioanniticus* and *C. israelensis*

Blindness in adults of the model cave fish *A. mexicanus* is a result of an embryonic process in which the rudimentary eye of the embryo is induced to degenerate by signals emitted from the lens tissue (Jeffery, 2009). Both early and late expression of RDGN genes, such as

Pax6, are responsible for the reduction of eyes in fish from cave populations (Jeffery, 2009; Strickler et al., 2001). Likewise, in the isopod crustacean *A. aquaticus* cave blindness has a strong genetic component and mechanisms of eye reduction also act at embryonic stages (Mojaddidi et al., 2018; Protas et al., 2011). The embryonic development of the reduced-eyes whip spider *C. israelensis* has not been explored to date, but we expect that reduction of eyes results from changes in embryonic gene expression during the deutembryo stage (Weygoldt, 1975). We investigated this possibility by quantifying the relative gene expression of RDGN genes in comparable embryonic stages of *C. israelensis* (reduced eyes) and *C. ioanniticus* (normal eyes) embryos before eye-spot formation (Additional file 1 Table S1; Fig. S1). Using the DGE approach from Comparison 1, we conducted a heterospecific analysis using as the reference either the *C. israelensis* transcriptome (henceforth "Comparison 2.1") or the *C. ioanniticus* transcriptome (henceforth "Comparison 2.2").

Both analyses are anchored on the premise that a hybrid mapping between the sister species is possible given their recent divergence. The mapping rate of the *C. ioanniticus* reads was similar regardless of the reference species, (96.74 and 96.59% respectively for *C. ioanniticus* and *C. israelensis*). In the case of the reads from *C. israelensis* embryos, mapping rate to the conspecific (96.8%) transcriptome was higher than when mapping against *C. ioanniticus* (82.45%). The similar mapping rate of *C. ioanniticus* reads suggests that the two whip spiders are sufficiently closely related to generate interspecific comparisons of gene expression. Comparisons 2.1 and 2.2 yielded similar results with respect to the direction of differentially expressed RDGN genes (Fig. 3b–c). Intriguingly, Comparison 2.1 shows that *Pax6A*, *OptixA* and *OptixB* are significantly more highly expressed in the reduced-eyes species, with expression levels at least 4 times higher than in the normal-eyes species (log2FC > 2; $p_{adj} < 0.05$) (Fig. 3b; Additional file 1, Fig. S10). In Comparison 2.2, *Pax6A* and *OptixA* are also more highly expressed in *C. israelensis* ($p_{adj} < 0.05$), and so is *eya* ($p_{adj} < 0.05$; Fig.

3c). In Comparison 2.2, *otd-B* appears more highly expressed in the normal-eyes species (p_{adj} < 0.05; Fig. 3c; Additional file 1, Fig. S11). We note that the magnitude of log2FC and significance values differed considerably between analyses. Nonetheless, *Pax6A* and *OptixA* were consistently more highly expressed in the reduced-eyes species, highlighting these two genes as promising candidates involved in the reduction of eyes in *C. israelensis*.

Expression of phototransduction genes and gene ontology enrichment analysis

Some of the RDGN genes surveyed are pleiotropic and expressed outside the eye field in other arthropods, including arachnids. For instance, *dac* is an important appendage-patterning gene (Kojima, 2004) and *otd* regulates anterior patterning across Arthropoda (Pechmann et al., 2009; Rosenberg et al., 2009). Therefore, our whole-embryo DGE comparisons may potentially not be sensitive enough to detect differences in expression in individual organs (i.e., eyes). In order to assess further the sensitivity of the approach to detecting eye-specific gene expression differences, we first quantified expression of opsins (visual pigments) and visual arrestins (phototransduction proteins) (Alvarez, 2008; Ramirez et al., 2016). We predicted that these retinal components should be upregulated in the eye-spot stages of *C. ioanniticus*, and expected that embryos of this eyed species should have higher expression of retinal genes in comparison to *C. israelensis*, the reduced-eyes species. We found three transcripts annotated as opsins for *C. ioanniticus* and *C. israelensis*: a *r-opsin* (Long-wave-sensitive clade 2 [LWS-2]), a *peropsin* and a *c-opsin* (Additional file 1 Fig. S12). LWS-2 (also referred to as Rh2) and *peropsins* are expressed on the eyes of some chelicerates, while *c-opsins* have been reported only in the central nervous system (Battelle et al., 2016; Eriksson et al., 2013; Nagata et al., 2012; Schomburg et al., 2015; Zopf et al., 2013).

For the visual arrestins, we recovered one homolog of *D. melanogaster Arrestin-2* (*Arr2*) in *C. ioanniticus*. We did not recover orthologs of *Arrestin-1* (*Arr1*) in either *Charinus*

transcriptome, but *Arr1* orthologs occur in the other chelicerate species surveyed (Additional file 1 Fig. S13). In addition, we discovered two *Arr2* paralogs in *C. ioanniticus* and *C. israelensis* that we termed *Arrestin-2- like* (*A/B*), given their close relationship to *Arr2 t*o the exclusion of *Arr1* and *D. melanogaster non-visual arrestin kutz* (Roman et al., 2000) (Additional file 1 Fig. S13).

In the intraspecific comparison between *C. ioanniticus* stages (Comparison 1), we detected that the *Arr2*, *c- opsin,* and *r-opsin* are significantly more highly expressed in the older stage with eye spots (Fig. 3d). In the interspecific Comparison 2.1, *c-opsin* and *r-opsin* are significantly more highly expressed in the eye bearing species (Fig. 3e). In Comparison 2.2, *Arr2, c-opsin*, and *r-opsin* are significantly more highly expressed in the eye bearing species (Fig. 3f). In this comparison, *Arrestin-2-like B* was more highly expressed in the reduced eye species, but we note that the identity of this arrestin needs to be further investigated, since it is does not cluster with the visual *Arr1* and *Arr2* (SI Appendix, Fig. S13). Taken together, these results suggest that our DGE approach is able to detect predicted differences in expression of downstream retinal genes between treatments.

Next, we conducted a Gene Ontology (GO) enrichment analysis in the gene sets of significantly highly and lowly expressed genes in the three comparisons, in order to investigate broader patterns of gene expression associated with eye development. We specifically looked for enrichment or depletion of the GO term “eye development” and child terms. In Comparison 1, we discovered enrichment of six eye-related GO terms only in the list of highly expressed genes in the stage before eyes form in *C. ioanniticus*, with 34 unique genes composing the enriched categories (Additional file 2, Table S3). These results accord with the detection of RDGN genes more highly expressed in the pre-eye stages (see above).

In Comparison 2.1, we detected ten eye-related enriched GO terms in the list of highly expressed genes in embryos of the blind species *C. israelensis*, with 187 unique genes

composing the enriched categories (Additional file 2, Table S3). The enrichment analysis of Comparison 2.2 yielded congruent results (seven eye-related GO terms; 125 unique genes) (Additional file 2, Table S3). The detection of eye-related GO terms enriched only on the more highly expressed genes of *C. israelensis* was unexpected, but is in accordance with the relative higher expression of *Pax6A* and *OptixA* detected in the analysis of RDGN genes (see above).

sine oculis is necessary for principal and secondary eye development in a model arachnopulmonate

Our bioinformatic analysis in the whip spider system suggested that *eya*, one paralog of *so*, and *otd* may be involved in the normal formation of eyes in *C. ioanniticus* (Comparison 1). We also found evidence that *Pax6* and a paralog of *Optix* may be involved in the reduction of eyes in the cave whip spider *C. israelensis*. To link bioinformatic measurements of gene expression with functional outcomes, we interrogated the function of RDGN genes using parental RNA interference (RNAi) in the spider *P. tepidariorum*. We selected *Ptep-soA* (*Ptep-so1* sensu Schomburg et al. 2015), *Ptep-otdB* (*Ptep-otd2* sensu Schomburg et al. 2015) and *Ptep-OptixB* (*Ptep Six3.2* sensu Schomburg et al. 2015). In *P. tepidariorum*, these genes are known to be expressed in all eye types, in the median eyes only, and in the lateral eyes, respectively (Fig. 1d) (Schomburg et al., 2015).

Early expression of *Ptep-soA* is detected in lateral domains of the head lobes (stage 10) corresponding to the principal and secondary eyes, and continues until the pre-hatching stage 14 (Schomburg et al., 2015). Expression of *Ptep-soA* on wild type stage 14.1 embryos is bilaterally symmetrical on all eyes and uniformly strong (Fig. 4a–b). By stage 14.2, it remains strong on the principal eyes but it is stronger at the periphery of the secondary eye spots (Fig. 4a, c).

Parasteatoda tepidariorum hatchlings, or postembryos, initially have no externally

visible lenses and pigment. The red pigment and lenses of all eyes, and the reflective tapetum of the lateral eyes, become progressively recognizable in the 48 h (at 26 °C) until the animal molts into the first instar with fully formed eyes (Additional file 3, Video S1) (see also Mittmann and Wolff, 2012). We fixed embryos from *Ptep-soA* dsRNA-injected and dH2O-injected treatments between 24 h–48 h, which encompasses stages where the eyes of postembryos are already recognizable until the first instar.

Negative control experiments (dH2O-injected females) yielded postembryos with eye morphology indistinguishable from wild type animals: the median eyes (ME; principal eyes) have an inferior semi-lunar ring of red pigment and lack the tapetum, and all pairs of lateral eyes (secondary eyes) have the canoe-shaped tapetum type (Foelix, 2011; Land, 1985), which is split in the middle and surrounded by red pigment (Fig. 5a; panel 1). We observed misshaped tapeta on the lateral eyes of some postembryos on the earlier side of the developmental spectrum of fixed animals, but that was never observed on postembryos close to molting or first instars (Additional file 1, Fig. S14). It is unclear if this reflects a natural variation of early developing tapetum or an artifact of sample preparation.

Embryos from *Ptep-soA* dsRNA-injected females are also able to hatch into postembryos and continue molting to adulthood (Additional file 3, Video S2). However, a subset of the embryos of dsRNA-injected treatment (9.5%; n = 195/2049) exhibits a spectrum of eye defects that was not observed on the controls (Fig. 5a–b; Additional file 1, Fig. S15). The defects occurred on all eyes, namely median eyes (ME), anterior lateral eyes (ALE), posterior lateral eyes (PLE), and median lateral eyes (MLE) (Fig. 5a). Affected median eyes have reduced pigmentation or complete absence (Fig. 5a, panels 2–6), while lateral eyes also exhibited defects of the tapetum or complete absence of the eye (Fig. 5a, panels 4–6).

We selected a subset of the knockdown postembryos initially scored as having any eye defect (n = 48) for quantifying the degree of effect per eye type, and the proportion of

symmetrical and mosaic eye phenotypes in our sample. Median eyes are affected in almost all cases (97%), whereas the three lateral eye types were similarly lowly affected (MLE: 14%; PLE: 8%; ALE: 10%) (Fig. 5c; Additional file 1, Fig. S14; detailed scoring criteria in Methods). The majority of defective eyes are mosaics, meaning that a given eye pair is affected only on one side of the animal (Fig. 5c; Additional file 1, Fig. S14).

Parental RNAi against *Ptep-soA* did not completely abolish its expression, as detected by in situ hybridization (Fig. 4d; see Methods). Nevertheless, we detected asymmetrical reduction of *Ptep-soA* expression on single eyes of a subset of stage 14 embryos (n = 6/16; Fig. 4d), which closely correlates with the predominance of mosaic phenotypes observed in late postembryos (Fig. 5c).

Parental RNAi experiments using the same protocol targeting *Ptep-otdB* and *Ptep-OptixB* did not result in any detectable phenotypic effects on the eyes of embryos from dsRNA-injected treatment (two and six females injected, respectively; counts not shown). These results accord with a recent study that knocked down both *Optix* paralogs *P. tepidariorum* and did not recover eye defects (Schacht et al., 2020).

Discussion

Duplication of RDGN members in arachnopulmonates Amblypygi have a critical placement within arachnid phylogeny, as they are part of a trio of arachnid orders (collectively, the Pedipalpi, comprised of Amblypygi, Thelyphonida, and Schizomida), which in turn is the sister group to spiders. Whereas the eyes of spiders have greatly diversified in structure, function, and degree of visual acuity (particularly the eyes of hunting and jumping spiders), the arrangement and number of eyes in Amblypygi likely reflects the ancestral condition across Tetrapulmonata (spiders + Pedipalpi), consisting of three pairs of simple lateral ocelli and a pair

of median ocelli; a similar condition is observed in basally branching spider groups like Mesothelae and Mygalomorphae, as well as Thelyphonida (vinegaroons). However, while developmental genetic datasets and diverse genomic resources are available for spiders and scorpions (Oda and Akiyama-Oda, 2020; Posnien et al., 2014; Schwager et al., 2017; Sharma et al., 2014a), the developmental biology of the other three arachnopulmonate orders has been virtually unexplored in the past four decades beyond the classic work describing the embryology of one North American amblypygid species (Weygoldt, 1975) (but see two recent studies on developmental patterning genes in Amblypygi [Gainett and Sharma, 2020; Harper et al., 2020]). To address this gap, we focused our investigation on a sister species pair of cave whip spiders and generated the first embryonic transcriptomes for this order. These datasets are immediately amenable to testing the incidence of RDGN duplicates previously known only from two spiders (Samadi et al., 2015; Schomburg et al., 2015) and their putative effects in patterning eyes across Arachnopulmonata broadly.

The inference of a partial or whole genome duplication (WGD) in the most recent common ancestor (MRCA) of Arachnopulmonata is supported by the systemic duplications of transcription factors and synteny detected in the genomes of the scorpion *Centruroides sculpturatus*, and the spider *P. tepidariorum*, as well as homeobox gene duplications detected in the genome of the scorpion *Mesobuthus martensii* and transcriptome of the spider *Pholcus phalangioides* (Leite et al., 2018; Schwager et al., 2017). Additional evidence comes from shared expression patterns of leg gap gene paralogs in a spider and a scorpion (Nolan et al., 2020). Embryonic transcriptomes are particularly helpful in the absence of genomes, as several duplicated genes, such as some homeobox genes, are only expressed during early stages of development (Leite et al., 2018; Sharma et al., 2014a; Sharma et al., 2015a). Our analysis of *Charinus* embryonic transcriptomes shows that RDGN gene duplicates observed in spiders also occur in whip spiders, supporting the hypothesis that these paralogs originated from a shared

WGD event in the common ancestor of Arachnopulmonata. This inference is independently corroborated by the occurrence of arachnopulmonate-specific Hox gene and leg gap gene duplicates in both *Charinus* and *Phrynus* transcriptomes (Gainett and Sharma, 2020) as well as the occurrence of *Wnt* and *frizzled* gene duplicates (Harper et al., 2020).

The conservation of some transcription factors patterning eyes is widespread in the metazoan tree of life (Vopalensky and Kozmik, 2009). In the model fruit fly *D. melanogaster*, the homeobox *Pax6* homolog eyeless was the first of several transcription factors identified as "master genes", necessary for compound eye formation and capable of inducing ectopic eye formation (Gehring and Ikeo, 1999; Kumar, 2009). The Pax6 protein is essential for eye formation across several metazoan taxa, which has fomented ample debate about the deep homology of gene regulatory networks in patterning structurally disparate eyes (Carroll, 2008; Shubin et al., 2009; Vopalensky and Kozmik, 2009). In the case of *so* (*Six1/2*), orthologs are found across metazoans (Bebenek et al., 2004; Byrne et al., 2018; Rivera et al., 2013). Evidence that *so* is required for the eye patterning in other bilaterians includes expression patterns in the developing eyes of the annelid *Platynereis dumerilii* (Arendt et al., 2002), and functional experiments in the planarian *Girardia tigrina* (Pineda et al., 2000). Therefore, studies interrogating the genetic bases of eye formation in chelicerate models have the potential to clarify which components of the eye gene regulatory network of Arthropoda evolved in the MRCA of the phylum, and which reflect deep homologies with other metazoan genes.

A conserved role for a *sine oculis* homolog in patterning arachnopulmonate eyes

The eyes of arthropods are diverse in number, arrangement, structure and function (Paulus, 1979). Both types of eyes observed in Arthropoda, the faceted eyes (compound) and single-lens eyes (ocelli), achieve complexity and visual acuity in various ways. To mention two extremes, in Mandibulata the compound eyes of mantis shrimps (Stomatopoda) achieve a

unique type of color vision and movements by using 12 different photoreceptive types and flexible eye-stalks (Daly et al., 2018; Marshall et al., 2007; Thoen et al., 2014). In Arachnida, the simple-lens median eyes of some jumping spiders (Salticidae) have exceptional visual acuity in relation to their eye size, achieve trichromatic vision through spectral filtering, and can move their retina using specialized muscles (Harland et al., 2012; Land, 1985; Zurek et al., 2015). Comparative anatomy suggests that the common ancestor of Arthropoda had both lateral compound eyes and median ocelli that then became independently modified in the arthropod subphyla (Morehouse et al., 2017; Paulus, 1979). In Chelicerata, the plesiomorphic eye condition is inferred to be a combination of median eyes (ocelli) and faceted eyes comparable to those of extant horseshoe crabs (Xiphosura), as well as extinct arachnid groups like Trigonotarbida and stem-group scorpions (e.g., *Proscorpius*) (Paulus, 1979). While in situ hybridization data for selected RDGN genes across arthropods generally support the hypotheses of eye homology, comparative developmental datasets remain phylogenetically sparse outside of Pancrustacea (Samadi et al., 2015; Schomburg et al., 2015).

We therefore applied a bioinformatic approach in a study system that lacked any genomic resources (Amblypygi) to assess whether RDGN homologs are transcriptionally active during the formation of eyes in the eye-bearing *C. ioanniticus*, as well as those that may be putatively involved in eye loss in its troglobitic sister species. As first steps toward understanding how arachnid eyes are patterned, our experiments demonstrated that *soA*, a *sine oculis* paralog identified as differentially expressed during the formation of eyes in *C. ioanniticus*, is necessary for patterning all eyes of a model arachnid system with the same eye configuration (*P. tepidariorum*). The reduction/loss of all eye types in the spider is consistent with the functional data in the beetle *T. castanaeum,* which demonstrates a role in compound eye formation (no ocelli occur in most beetles) (Yang et al., 2009), and in *D. melanogaster* and the cricket *G. bimaculatus*, in which both compound eyes and ocelli are affected (Kumar, 2009;

Takagi et al., 2012). Thus, we provide the first functional evidence that part of the RDGN is evolutionarily conserved in the MRCA of insects and arachnids and by extension, across Arthropoda.

The advantage of such a bioinformatic approach is that it can potentially narrow the range of candidate genes for functional screens, due to the inherent challenges imposed by duplications when assessing gene function. Eye reduction in the cave fish *A. mexicanus* has been shown to involve differential expression of genes known to be involved in eye patterning in model organisms, such as *hedgehog* and *Pax6* (Jeffery, 2009; Protas and Jeffery, 2012). In addition, other "non-traditional" candidates have been identified, such as *hsp90* (Jeffery, 2009). Likewise, evidence from quantitative trait loci mapping in cave populations of the troglobitic crustacean *A. aquaticus* shows that eye loss phenotype is correlated with loci that are not part of the RDGN (Protas and Jeffery, 2012; Protas et al., 2011). The results of the DGE analysis in whip spiders underscore the potential of a DGE approach to triangulate targets among candidate genes in non-model species more broadly. Future efforts in the *Charinus* system should focus on dissecting individual eye and limb primordia of embryos of both species, in order to identify candidate genes putatively involved in the reduction of each eye type, as well as compensatory elongation of the sensory legs of the troglobitic species, toward down- stream functional investigation.

Do gene duplications play a role in the functional diversification of arachnopulmonate eyes?

A challenge in studying arachnopulmonate models to understand ancestral modes of eye patterning in Arthropoda is the occurrence of RDGN duplicates in this lineage. Our orthology searches and phylogenetic analysis showed that the evolutionary history of genes is not always resolved using standard phylogenetic methods, as short alignable regions and/or uncertainty of multiple sequence alignments can result in ambiguous gene trees. One way to

circumvent this limitation is by analyzing expression patterns via in situ hybridization between paralogs in different arachnids in order to determine which patterns are plesiomorphic (Leite et al., 2018; Nolan et al., 2020; Turetzek et al., 2015). Nonetheless, the possibility of subfunctionalization and neofunctionalization may also complicate such inferences because discerning one process from the other is analytically challenging (Sandve et al., 2018).

Genetic compensation of gene paralogs is another confounding variable; indeed, at least the redundancy of *Pax6* paralogs is inferred to be ancient in arthropods (Friedrich, 2017). Deciphering the potentially overlapping or redundant functions of paralogs can be accounted for by experimental advances in model organisms (e.g., [Shull et al., 2020]), but comparable advances can be challenging for new systems. We note that the overall penetrance in our experiment is low (9.5%) when compared to some studies in *P. tepidariorum* (e.g., [(Khadjeh et al., 2012)]; > 59% in *Ptep-Antp* RNAi). Wide variance in penetrance has been reported by several research groups in this system, with phenotypic effects varying broadly even within individual experiments (e.g., Fig. 5 of [Akiyama-Oda and Oda, 2006]; Fig. S5 of [Schwager et al., 2009]). Furthermore, some genes have empirically proven intractable to transcript degradation by RNAi in *P. tepidariorum*, with one case suggesting functional redundancy to be the cause (posterior Hox genes [Khadjeh et al., 2012]). Double knockdown experiments have been shown to exhibit poor penetrance (0–1.5%) in *P. tepidariorum* as well (Fig. S3 of [Khadjeh et al., 2012]; Fig. S1 of [Setton et al., 2017]), and to our knowledge, no triple knockdown has ever been achieved. While we cannot rule out functional redundancy with other RDGN paralogs in the present study, the low penetrance we observed may also be partly attributable to our conservative phenotyping strategy (see Methods), which did not assess a possible delay in eye formation and emphasized dramatic defects in eye morphology for scoring.

The occurrence of RDGN gene duplications in Arachnopulmonata, in tandem with improving functional genetic toolkits in *P. tepidariorum* (e.g., (Pechmann, 2016)), offers a

unique opportunity for studying the role of sub- and neofunctionalization during the development of their eyes, and a possible role for these processes in the diversification of number, position and structure of the eyes in an ancient group of arthropods (Harland et al., 2012; Land, 1985; Morehouse et al., 2017; Paulus, 1979; Zurek et al., 2015). Genomic resources for mites, ticks, and harvestmen (Grbić et al., 2011; Gulia-Nuss et al., 2016; Hoy et al., 2016) reveal that apulmonate arachnid orders have not undergone the genome duplication events exhibited by Arachnopulmonata (Schwager et al., 2017) and separately by horseshoe crabs (Kenny et al., 2015; Nossa et al., 2014; Zhou et al., 2020). Future comparative studies focused on understanding the ancestral role of chelicerate RDGN genes should additionally prioritize single-copy orthologs in emerging model systems independent of the arachnopulmonate gene expansion, such as the harvestman *Phalangium opilio* (Sharma et al., 2012; Sharma et al., 2013).

Conclusions

Our work establishes a foundation to pursue the genetics of eye loss in cave arachnids, both by establishing a whip spider study system for comparative investigation, and by linking differential gene expression to an arthropod eye phenotype for the first time outside of Pancrustacea. Considering the phylogenetic position of arachnids, this finding implies that at least one of the classic eye genes discovered in insect model species had a conserved function in the common ancestor of Arthropoda. The systemic gene duplications in these arachnids offer a promising system for investigating the role of ohnologs in the diversification of arachnid eyes.

Methods

Animal collection

Three ovigerous females of the normal-eyes species, *C. ioanniticus* (ISR021–2; ISR021–3; ISR021–4), and two egg-carrying females of the reduced-eyes species, *C. israelensis* (ISR051–4; ISR051–6), were hand collected in caves in Israel in August 2018 (Supplementary Information; Table 1). Females were sacrificed and the brood sacs containing the embryos were dissected under phosphate saline buffer (PBS). For each female, a subset of the embryos (5 to 13 individuals) was fixed in RNAlater (ThermoFisher) after poking a whole into the egg membrane with fine forceps, while the remaining embryos of the clutch were fixed in a 4% formaldehyde/PBS solution to serve as vouchers (Additional file 1, Table S1). Adult animals and embryos of *P. tepidariorum* were obtained from the colony at UW-Madison, US, in turn derived from a laboratory culture founded with spiders collected near Cologne, Germany (Schwager et al., 2015).

Transcriptome assembly for *Charinus* whip spiders

RNAlater-fixed embryos were transferred to 1.5 mL tubes filled with TRIZOL (Invitrogen) after 2 months, and subject to RNA extraction. Total RNA extracted from each sample of the embryos of *C. ioanniticus* (three samples) and *C. israelensis* (two samples) (Additional file 1, Table S1) was submitted for library preparation at the Biotechnology Center of the University of Wisconsin - Madison. Each sample was sequenced in triplicate in an Illumina High-Seq platform using paired-end 100 bp-long read strategy at the same facility. Read quality was assessed with FastQC (Babraham Bioinformatics). Paired-end reads for *C. ioanniticus* (ISR021) and *C. israelensis* (ISR051) were compiled and de novo assembled using Trinity v.3.3 (Grabherr et al., 2011) enabling Trimmomatic v.0.36 to remove adapters and low-quality reads (Bolger et al., 2014). Transcriptome quality was assessed with the Trinity package script ‘TrinityStats.pl’ and BUSCO v.3 (Waterhouse et al., 2017). For BUSCO, we used the

'Arthropoda' database and analyzed the transcriptomes filtered for the longest isoform per Trinity gene.

RNA sequencing for differential gene expression

The total RNA extraction of each sample of *C. ioanniticus* and *C. israelensis* embryos was sequenced in triplicate in an Illumina High-Seq platform using a single-end 100 bp-long read strategy in the same facility as described above. For *C. ioanniticus* (normal-eyes), we sequenced two biological replicates of embryos at an early embryonic stage, before eye-spot formation (ISR021–2, ISR021–3), and one sample of late embryos, after eye-spot formation (ISR021–4). For *C. israelensis* (reduced-eyes), we sequenced embryos at an early embryonic stage (ISR051–6; ISR051–4) comparable to the early stage in *C. ioanniticus* (ISR021–2, ISR021–3), as inferred by the elongated lateral profile of the body and marked furrows on the opisthosomal segments (Additional file 1, Fig. S1).

Differential gene expression analysis in Charinus and identification of eye gene orthologs

Orthologs of *D. melanogaster ey* and *twin of eyeless* (*Pax6A, Pax6B*)*, sine oculis* (*soA, soB*)*, orthodenticle* (*otdA, otdB*)*, Optix* (*Six3.1, Six3.2*)*, dachshund* (*dacA, dacB*)*, and eyes absent* (*eya*) had been previously isolated in *P. tepidariorum* (Schomburg et al., 2015, and references therein). We used as reference sequences the complete predicted transcripts for these genes from *P. tepidariorum* genome (Schwager et al., 2017), *Cupiennius salei* (Samadi et al., 2015) (for *atonal [ato]* and *Pax6*), and *D. melanogaster*, including also *ato* and *eyegone* (*eyg*) from the latter species. The sequences were aligned with MAFFT (v7.407) (Katoh and Standley, 2013) and the resulting alignments were used to build hidden Markov model profiles for each gene (hmmbuild, from the hmmer suite v.3.3) (Finn et al., 2015). Matches to these profiles were

found using hmmsearch in the reference transcriptomes of *C. ioanniticus* and *C. israelensis* as well as in the genomes of representative arthropods, including *D.melanogaster* (GCA 000001215.4), *T. castaneum* (GCA 000002335.3), *Daphnia magna* (GCA 003990815.1), *Strigamia maritima* (GCA 000239455.1), *Dinothrombium tinctorium* (GCA 003675995.1), *Ixodes scapularis* (GCA 002892825.2), *Tetranychus urticae* (GCA 000239435.1), *Limulus polyphemus* (GCA 000517525.1), *Tachypleus tridentatus* (GCA 004210375.1), *C. sculpturatus* (GCA 000671375.2), *P. tepidariorum* (GCA 000365465.2) and *Trichonephila clavipes* (GCA 002102615.1). These species were selected from a pool relatively recent genome assembly resources and well curated reference genomes.

Homologous sequences (those with hmmer expectation value, e < 1010) to the genes of interest were then compiled into individual gene FASTA files, combined with the reference sequences used for the homology search, aligned (MAFFT [Katoh and Standley, 2013]), trimmed of gap rich regions (trimAL v.1.2, −gappyout) (Capella-Gutiérrez et al., 2009) and used for maximum likelihood gene tree estimation (IQ-TREE v.1.6.8, −mset LG,WAG, JTT,DCMUT –bb 1000) (Nguyen et al., 2015). The association of transcripts in the *Charinus* species with the genes of interest is based on the gene phylogeny and was followed by inspection of the coding sequences to distinguish splicing variants from other gene paralogs. Alignments, newick trees, and the list of *Charinus* sequences are available in Additional file 4 Dataset S1. The gene transcript association was then used to generate the transcript-to-gene map required for the DGE analysis.

For the analysis of opsins, protein sequences of the five *P. tepidariorum* opsins identified in a previous study (Schomburg et al., 2015) were used as queries for tblastn searches, and candidates were reciprocally blasted against NCBI non-redundant sequence database. For identification of arrestin homologs, the same procedure was performed using *D. melanogaster Arr1* (FBpp0080583) and *Arr2* (FBpp0076326) as queries. Protein sequences of metazoan

opsins from previous studies (Battelle et al., 2016; Morehouse et al., 2017) were aligned with candidate opsins (MAFFT [Katoh and Standley, 2013]), and gene trees were inferred in a maximum likelihood phylogenetic analysis (IQ-TREE v.1.6.8, −m TEST –bb 1000). The annotation of arrestins was based on the nomenclature and reference protein sequences in (Alvarez, 2008), supplemented with arrestins identified by blastp in the genomes of *T. castaneum*, *L. polyphemus*, *P. tepidariorum*, *C. sculpturatus*, *I. scapularis* and *T. urticae*. A *Homo sapiens* alfa arrestin (NP_056498.1) was used as a reference outgroup. Alignment and gene tree inference were performed as above.

Read mapping, transcript abundance quantification, and GO enrichment analysis

For the in silico analysis of gene expression, single-end raw reads were first trimmed using the software Trimmomatic v. 0.35 (Bolger et al., 2014). For the intraspecific analysis of early (before eyespot) and late (eyespot) embryos of *C. ioanniticus* (Comparison 1), the trimmed reads were quantified in the embryonic transcriptome of *C. ioanniticus*. For the interspecific comparison of early embryos of *C. ioanniticus* and *C. israelensis*, two reciprocal analysis were conducted: reads from both species mapped onto *C. israelensis* transcriptome as the reference (Comparison 2.1); and reads from both species mapped onto *C. ioanniticus* transcriptome (Comparison 2.2).

Transcript abundance was quantified using the software Salmon v. 1.1.0 (Patro et al., 2017), enabling the flag '–validate- Mapping'. Analysis of differential gene expression was conducted with the software DESeq2 v 1.24.0 (Love et al., 2014) following a pipeline with the R package tximport v.1.12.3 (Soneson et al., 2015). The exact procedures are documented in the custom R script (Additional file 5, Dataset S2).

For the enrichment analysis, we annotated both transcriptomes using the Trinotate

v.3.2.1 pipeline (Bryant et al., 2017) and extracted GO term annotations with ancestral GO terms using the package script extract_GO_assignments _from_ Trinotate_xls.pl. We conducted the GO enrichment analysis using the R package Goseq v.1.40.0 (Young et al., 2010), as implemented by a modified Trinity v.2.8.5 script run_ GOseq.pl (Haas et al., 2013). Enrichment analyses were conducted for Comparison 1, Comparison 2.1 and Comparison 2.2, separately for the upregulated (log2FC > 1) and down- regulated (log2FC < 1) set of significant genes ($p_{adj} \leq 0.05$). We considered a GO term enrichment or depleted if FDR ≤ 0.05 (Additional file 6, Dataset S3). We searched for enriched GO terms associated with eye development (GO:0001654), and daughter GO terms (177 GO identifiers) as retrieved by the function get_ child_nodes in R package GOfuncR v.1.8.0 (Grote, 2020).

Parental RNA interference, in situ hybridization, and imaging in *Parasteatoda tepidariorum*

Total RNA from a range of embryonic stages of *P. tepidariorum* was extracted with TRIZOL (Invitrogen), and cDNA was synthetized using SuperScriptIII (Invitrogen). Gene fragments for *Ptep-soA, Ptep-otdB*, and *Ptep- OptixB* were amplified from cDNA using gene specific primers designed with Primers3Web version 4.1.0 (Koressaar and Remm, 2007) and appended with T7 ends. Cloning amplicons were generated using the TOPO TA Cloning Kit with One Shot Top10 chemically competent *Escherichia coli* (Invitrogen). Amplicon identities and directionality were assessed with Sanger sequencing. Primer, amplicon sequences and fragment lengths are available in Additional file 7 Dataset S4. Double-stranded RNA for *Ptep-soA, Ptep-otdB* and *Ptep-OptixB* was synthesized using the MEGAScript T7 transcription kit (Thermo Fisher). Sense and antisense RNA probes for colorimetric in situ hybridization were synthesized from plasmid templates with DIG RNA labeling mix (Roche) and T7/T3 RNA polymerase (New England Biolabs).

Parental RNA interference (RNAi) followed established protocols for double-stranded RNA (dsRNA) injection in virgin females of *P. tepidariorum* (Oda and Akiyama-Oda, 2020). Each female was injected four times with 2.5 μL of dsRNA at a concentration of 2 μg/uL, to a total of 20 μg. For *Ptep-soA*, seven virgin females were injected with dsRNA of a 1048 bp cloned fragment (Additional file 1, Fig. S15C) and 3 females were injected with the same volume of dH2O as a procedural control. Two virgin females were injected with dsRNA for *Ptep-otdB*, and six females for *Ptep-OptixB*. All females were mated after the second injection and were fed approximately every-other day after the last injection. Cocoons were collected until the sixth clutch, approximately once per week.

Hatchlings for all cocoons were fixed between 24 and 48 h after hatching. Freshly hatched postembryos have almost no external signs of eye lenses and pigments. The selected fixation window encompasses a period in which postembryos have deposited eye pigments until the beginning of the first instar, where eyes are completely formed (Additional file 3, Video S1, S2). Hatchlings were immersed in 25% ethanol/PBST and stored at 4 °C. For the *Ptep-soA* RNAi experiment, hatchlings were scored in four classes: (1) wild type, where all eyes were present and bilaterally symmetrical; (2) eyes defective, where one or more eyes were reduced in size or completely absent; (3) dead/arrested; (4) undetermined, where embryos were damaged or clearly freshly hatched. A subset of *Ptep-soA* dsRNA-injected embryos from four clutches (n = 48) and of three control clutches (n = 48) were further inspected in detail to assess the effects on individual eye types. Given that there is a spectrum on the intensity of pigment deposition in the median eyes (ME), and small asymmetries on the shape of the early developing tapetum of the lateral eyes (LE) in control embryos, the following conservative criteria were adopted: (1) ME were considered affected when asymmetry in pigmentation or lens size was detected; both ME were only scored as affected when they were both completely missing, in order to rule out embryos that were simply delayed in pigment deposition; (2) LE were

considered defective only when the tapetum was completely absent (Additional file 1, Fig. S14). Therefore, our coding does not allow detection of a phenotype consisting of delayed pigmentation. Raw data are available in Additional file 7 Dataset S4.

For in situ hybridization, a subset of *Ptep-soA* dsRNA- injected embryos at stage 13/14 (Mittmann and Wolff, 2012) was fixed in a phase of heptane and 4% formaldehyde for 12–24 h, washed in PBST, gradually dehydrated in methanol and stored at − 20 °C for at least 3 days before downstream procedures, after a modified protocol of Akiyama-Oda and Oda (2003). In situ hybridization followed the protocol of Akiyama-Oda and Oda (2003).

Embryos from in situ hybridization were counterstained with Hoechst 33342 and imaged using a Nikon SMZ25 fluorescence stereomicroscope mounted with a DS- Fi2 digital color camera (Nikon Elements software). For postembryos, the prosoma was dissected with fine forceps, gradually immersed in 70% Glycerol/PBS-T and mounted on glass slides. Postembryos were imaged using an Olympus DP70 color camera mounted on an Olympus BX60 epifluorescence compound microscope.

The datasets in Additional files 3, 4, 5, 6 and 7 are deposited in: https://datadryad.org/stash/share/xb2VW4o8 0AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM

doi:https://doi.org/10.5061/dryad.xgxd254d1

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-020-07149-x.

Additional file 1. Figs. S1–S15 and Tables S1–S2. (.pdf)

Additional file 2. Table S3: Subset of enriched GO terms that have eye- related ontology for Comparison 1, Comparison 2.1, and Comparison 2.2. Each spreadsheet is

accompanied by the Trinotate annotation report of the differentially expressed genes in the enriched eye-related GO terms. GO enrichment analyses absent from this file do not have eye-related GO terms enriched. For the full GOseq results for each comparison see Additional file 6, Dataset S3. (.xlsx).

Additional file 3: Video S1: Time-lapse imaging of a postembryo ~ 24 h after hatching of *Parasteatoda tepidariorum* from the dH2O-injected treatment (negative control). Pictures were taken every 30 min, in a room at 22 °C. Normal molting time after hatching is ~ 48 h at 26 °C. Video S2: Time-lapse imaging of a postembryo ~ 24 h after hatching of *Parasteatoda tepidariorum* from the *Ptep-soA*-injected treatment. Pictures were taken every 30 min, in a room at 22 °C. Normal molting time after hatching is ~ 48 h at 26 °C. (.zip). Available at: https://datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM; doi:https://doi.org/ 10.5061/dryad.xgxd254d1

Additional file 4. Dataset S1: Dataset of the orthology analyses. Alignments and sequences of *Charinus* RDGN genes identified in this study and gene trees in Newick format; alignments of opsins and arrestins. (.zip). Available at: https://datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM; doi:https://doi.org/10.5061/dryad.xgxd254d1

Additional file 5. Dataset S2: Dataset for the differential gene expression analyses. DESeq2 dataset (DESeq (dds); filtered for $p_{adj} > 0.5$) of the DGE analysis of Comparison 1, 2.1 and 2.2 (see "Material and Methods" for explanation). run_ddseq2.r: custom R script used to run all three analysis. (.zip). Available at: https://datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM; doi:https://doi.org/10.5061/dryad.xgxd254d1

Additional file 6. Dataset S3: Dataset for the GOseq enrichment analyses. This folder contains the complete GOseq results of enriched/depleted (FDR ≤ 0.05) GO terms of up-

(Log2FC > 1) and down-regulated (Log2FC < 1) genes in Comparisons 1, 2.1 and 2.2. The twelve spreadsheets (.xls) are named as follows: Comparison#,_UP/DOWN_enriched/depleted. (.zip). Available at: https:// datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2 miM; doi:https://doi.org/10.5061/dryad.xgxd254d1

Additional file 7. Dataset S4. (a): High resolution Additional file 1 Fig. S14. (b): Spreadsheets with the raw counts and sum of counts used to generate the distribution bar plots of the *Ptep-soA* RNAi experiment. (c) raw counts and sum of counts used to generate the distribution bar plots of the effects of *Ptep-soA* RNAi per eye type. (d) Primer sequences for the amplified fragments of *Ptep-soA, Ptep-otdB* and *Ptep-OptixB*. (.zip). Available at: https://datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM; doi:https://doi.org/10.5061/dryad. xgxd254d1

Abbreviations

atoA/B: atonalA/atonalB; *dacA/B: dachshundA/B*; *eya: eyes absent*; *eyg: eyegone*; *otdA/B: orthodenticleA/B*; RDGN: Retinal Determination Gene Network; *soA/B: sine oculisA/B*; ME: Median eyes; ALE: Anterior lateral eyes; PLE: Posterior lateral eyes; MLE: Median lateral eyes; LE: Lateral eyes

Acknowledgements

Microscopy was performed at the Newcomb Imaging Center, Department of Botany, University of Wisconsin-Madison. Sequencing was performed at the UW-Madison Biotechnology Center. Access to computing nodes for intensive tasks was provided by the Center for High Throughput Computing (CHTC) and the Bioinformatics Resource Center

(BRC) of the University of Wisconsin– Madison. Comments from three anonymous reviewers substantially improved an earlier version of the manuscript.

Authors' contributions

J.A.B, E.G.R., and P.P.S. designed the study. G.G., J.A.B., S.A., E.G.R., and PPS conducted fieldwork. G. G., J.A.B., C.R.K., J.T.Z., J.M.Z., S.A., and P.P.S. performed RNAi experiments and fixed specimens. G.G. and P.P.S. analyzed data from RNAi experiments. G.G., J.A.B., and P.P.S. extracted RNA and assembled the *Charinus* embryonic transcriptomes. G.G. and P.P.S. conducted in situ hybridizations. G.G. conducted GO enrichment analyses. J.A.B. conducted orthology searches and analysis of differential gene expression. J.A.B., E.G.R., and P.P.S. obtained funding. G.G., J.A.B., and P.P.S. produced a first draft of the manuscript. All authors revised the final version of the manuscript.

Funding

Fieldwork in Israel was supported by a National Geographic Society Expeditions Council grant no. NGS-271R-18 to J.A.B. Sequencing, reagents for molecular work, and access to computer clusters were funded by National Science Foundation grant no. IOS-1552610 and IOS-2016141 to P.P.S. G.G. was supported by a Wisconsin Alumni Research Foundation Fall Research Competition award. Funding agencies played no part in the design of the study, data collection, data analysis, data interpretation, or drafting of the manuscript.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Additional files 2–7 Videos and Datasets in the Dryad repository

(https://datadryad.org/stash/share/xb2VW4o80AmId3mLFZB07Ho7rHxPx-htK9q5J_-2miM). Raw reads are deposited in NCBI SRA database under accession number PRJNA649577.

Ethics approval and consent to participate

Specimens were collected under permit 2018/42037, issued by the Israel National Parks Authority to E.G.R.

Consent for publication

Not applicable.

Competing interests

The authors declare having no conflict of interest.

References

Akiyama-Oda, Y. and Oda, H. (2003). Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* 130, 1735–1747.

Akiyama-Oda, Y. and Oda, H. (2006). Axis specification in the spider embryo: *dpp* is required for radial-to-axial symmetry transformation and *sog* for ventral patterning. *Development* 133, 2347–2357.

Alvarez, C. E. (2008). On the origins of arrestin and rhodopsin. *BMC Evol. Biol.* 8, 222–13.

Arendt, D., Tessmar, K., de Campos-Baptista, M.-I. M., Dorresteijn, A. and Wittbrodt, J. (2002). Development of pigment-cup eyes in the polychaete *Platynereis dumerilii* and evolutionary conservation of larval eyes in Bilateria. *Development* 129, 1143–1154.

Ballesteros, J. A. and Sharma, P. P. (2019). A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* 33, 440–22.

Ballesteros, J. A., López, C. E. S., Kováč, Ľ., Gavish-Regev, E. and Sharma, P. P. (2019). Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. *Proc. R. Soc. B* 286, 20192426.

Barnett, A. A. and Thomas, R. H. (2012). The delineation of the fourth walking leg segment is temporally linked to posterior segmentation in the mite *Archegozetes longisetosus* (Acari: Oribatida, Trhypochthoniidae). *Evol. Dev.* 14, 383–392.

Barnett, A. A. and Thomas, R. H. (2013a). Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *Evodevo* 4, 23.

Barnett, A. A. and Thomas, R. H. (2013b). The expression of limb gap genes in the mite *Archegozetes longisetosus* reveals differential patterning mechanisms in chelicerates. *Evol. Dev.* 15, 280–292.

Battelle, B.-A., Ryan, J. F., Kempler, K. E., Saraf, S. R., Marten, C. E., Warren, W. C., Minx, P. J., Montague, M. J., Green, P. J., Schmidt, S. A., et al. (2016). Opsin repertoire and expression patterns in horseshoe crabs: Evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biol. Evol.* 8, 1571–1589.

Bebenek, I. G., Gates, R. D., Morris, J., Hartenstein, V. and Jacobs, D. K. (2004). *sine oculis* in basal Metazoa. *Dev. Genes Evol.* 214, 342–351.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Bradic, M., Teotónio, H. and Borowsky, R. L. (2013). The Population Genomics of Repeated Evolution in the Blind Cavefish *Astyanax mexicanus*. *Mol. Biol. Evol.* 30, 2383–2400.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776.

Byrne, M., Koop, D., Morris, V. B., Chui, J., Wray, G. A. and Cisternas, P. (2018). Expression of genes and proteins of the pax-six-eya-dach network in the metamorphic sea urchin: Insights into development of the enigmatic echinoderm body plan and sensory structures. *Dev. Dyn.* 247, 239–249.

Cagan, R. (2009). Chapter 5 - Principles of *Drosophila* Eye Differentiation. In *Current Topics in Developmental Biology*, pp. 115–135. Current Topics in Developmental Biology.

Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. (2009). trimAl: a tool for

automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* 134, 25–36.
Coghill, L. M., Hulsey, C. D., Chaves-Campos, J., Leon, F. J. G. de and Johnson, S. G. (2014). Next generation phylogeography of cave and surface *Astyanax mexicanus*. *Mol. Phylogenet. Evol.* 79, 368–374.
Cruz-López, J. A., Proud, D. N. and Pérez-González, A. (2016). When troglomorphism dupes taxonomists: morphology and molecules reveal the first pyramidopid harvestman (Arachnida, Opiliones, Pyramidopidae) from the New World. *Zool. J. Linn. Soc.* 177, 602–620.
Daly, I. M., How, M. J., Partridge, J. C. and Roberts, N. W. (2018). Complex gaze stabilization in mantis shrimp. *Proc. R. Soc. B* 285, 20180594.
Derkarabetian, S., Steinmann, D. B. and Hedin, M. (2010). Repeated and time-correlated morphological convergence in cave-dwelling harvestmen (Opiliones, Laniatores) from montane Western North America. *PLoS One* 5, e10388.
Eriksson, B. J., Fredman, D., Steiner, G. and Schmid, A. (2013). Characterization and localization of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. *BMC Evol. Biol.* 13, 186.
Esposito, L. A., Bloom, T., Caicedo-Quiroga, L., Alicea-Serrano, A. M., Sánchez-Ruíz, J. A., May-Collado, L. J., Binford, G. J. and Agnarsson, I. (2015). Islands within islands: Diversification of tailless whip spiders (Amblypygi, *Phrynus*) in Caribbean caves. *Mol. Phylogenet. Evol.* 93, 107–117.
Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A. and Eddy, S. R. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–8.
Foelix, R. (2011). *Biology of spiders*. 3rd ed. OUP USA.
Friedrich, M. (2017). Ancient genetic redundancy of eyeless and twin of eyeless in the arthropod ocular segment. *Dev. Biol.* 432, 192–200.
Gainett, G. and Sharma, P. P. (2020). Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *Evodevo* 11, 18.
Garwood, R. J., Sharma, P. P., Dunlop, J. A. and Giribet, G. (2014). A Paleozoic stem group to mite harvestmen revealed through integration of phylogenetics and development. *Curr. Biol.* 24, 1017–1023.
Gehring, W. J. and Ikeo, K. (1999). Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet.* 15, 371–377.
Giribet, G. (2018). Current views on chelicerate phylogeny—A tribute to Peter Weygoldt. *Zool. Anz. J. Comp. Zool.* 273, 7–13.
Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
Grbić, M., Khila, A., Lee, K.-Z., Bjelica, A., Grbić, V., Whistlecraft, J., Verdon, L., Navajas, M. and Nagy, L. (2007). Mity model: *Tetranychus urticae*, a candidate for chelicerate model organism. *BioEssays* 29, 489–496.
Grbić, M., Leeuwen, T. Van, Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., et al. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479, 487–492.
Grote, S. (2020). *GOfuncR: Gene ontology enrichment using FUNC. R package version*

1.8.0.

Gulia-Nuss, M., Nuss, A. B., Meyer, J. M., Sonenshine, D. E., Roe, R. M., Waterhouse, R. M., Sattelle, D. B., Fuente, J. D. La, Ribeiro, J. M., Megy, K., et al. (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* 7, 10507.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.

Harland, D. P., Li, D. and Jackson, R. R. (2012). How jumping spiders see the world. In *How Animals See the World: Comparative Behavior, Biology, and Evolution of Vision* (ed. Lazareva, O. F., Shimizu, T., and Wasserman, E. A.), pp. 132–163. Oxford Academic.

Harper, A., Baudouin-Gonzalez, L., Schönauer, A., Seiter, M., Holzem, M., Arif, S., McGregor, A. P. and Sumner-Rooney, L. (2020). Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates. *bioRxiv* 160, 1067–40.

Harvey, M. S. (2002). The neglected cousins: What do we know about the smaller arachnid orders? *J. Arachnol.* 30, 357–372.

Harvey, M. S. (2007). The smaller arachnid orders: Diversity, descriptions and distributions from Linnaeus to the present (1758 to 2007). *Zootaxa* 1668, 363–380.

Hedin, M. and Thomas, S. M. (2010). Molecular systematics of eastern North American Phalangodidae (Arachnida: Opiliones: Laniatores), demonstrating convergent morphological evolution in caves. *Mol. Phylogenet. Evol.* 54, 107–121.

Herman, A., Brandvain, Y., Weagley, J., Jeffery, W. R., Keene, A. C., Kono, T. J. Y., Bilandžija, H., Borowsky, R., Espinasa, L., O'Quin, K., et al. (2018). The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. *Mol. Ecol.* 27, 4397–4416.

Homann, H. (1971). Die Augen der Araneae. *Z. Morph. Tiere.* 69, 201–272.

Howarth, F. G. (1993). High-stress subterranean habitats and evolutionary change in cave-inhabiting arthropods. *Am. Nat.* 142 Suppl 1, S65–77.

Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., Pomerantz, A. F., Simão, F. A., Thomas, J., Jiggins, F. M., et al. (2016). Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomized Hox genes and superdynamic intron evolution. *Genome Biol. Evol.* 8, 1762–1775.

Jeffery, W. R. (2009). Chapter 8. Evolution and development in the cavefish *Astyanax*. *Curr. Top. Dev. Biol.* 86, 191–221.

Jemec, A., Škufca, D., Prevorčnik, S., Fišer, Ž. and Zidar, P. (2017). Comparative study of acetylcholinesterase and glutathione S-transferase activities of closely related cave and surface *Asellus aquaticus* (Isopoda: Crustacea). *PLoS One* 12, e0176746–14.

Juan, C., Guzik, M. T., Jaume, D. and Cooper, S. J. B. (2010). Evolution in caves: Darwin's "wrecks of ancient life" in the molecular era. *Mol. Ecol.* 19, 3865–3880.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., et al. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity* 116, 190–199.

Khadjeh, S., Turetzek, N., Pechmann, M., Schwager, E. E., Wimmer, E. A., Damen, W. G. M. and Prpic, N. M. (2012). Divergent role of the Hox gene *Antennapedia* in spiders

is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl Acad. Sci. USA* 109, 4921–4926.
Kojima, T. (2004). The mechanism of *Drosophila* leg development along the proximodistal axis. *Dev. Growth Differ.* 46, 115–129.
Koressaar, T. and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.
Kumar, J. P. (2009). The molecular circuitry governing retinal determination. *Biochim. Biophys. Acta* 1789, 306–314.
Land, M. F. (1985). The Morphology and Optics of Spider Eyes. In *Neurobiology of Arachnids*, pp. 53–78. Springer-Verlag Berlin Heidelberg.
Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N.-M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018). Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* 35, 2240–2253.
López, C. E. S., Francke, O. F. and Prendini, L. (2014). Shining a light into the world's deepest caves: phylogenetic systematics of the troglobiotic scorpion genus *Alacran* Francke, 1982 (Typhlochactidae:Alacraninae). *Invertebr. Syst.* 28, 643–664.
Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15, 550.
Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D. and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* 10, 2295–8.
Mammola, S. and Isaia, M. (2017). Spiders in caves. *Proc. R. Soc. B* 284, 20170193.
Mammola, S., Mazzuca, P., Pantini, P., Isaia, M. and Arnedo, M. A. (2017). Advances in the systematics of the spider genus *Troglohyphantes* (Araneae, Linyphiidae). *Syst. Biodivers.* 15, 307–326.
Mammola, S., Cardoso, P., Ribera, C., Pavlek, M. and Isaia, M. (2018a). A synthesis on cave-dwelling spiders in Europe. *J. Zool. Syst. Evol. Res.* 56, 301–316.
Mammola, S., Arnedo, M. A., Pantini, P., Piano, E., Chiappetta, N. and Isaia, M. (2018b). Ecological speciation in darkness? Spatial niche partitioning in sibling subterranean spiders (Araneae: Linyphiidae: Troglohyphantes). *Invert. Systematics* 32, 1069–1082.
Marshall, J., Cronin, T. W. and Kleinlogel, S. (2007). Stomatopod eye structure and function: A review. *Arthropod Struct. Dev.* 36, 420–448.
Miranda, G. S., Aharon, S., Gavish-Regev, E., Giupponi, A. P. L. and Wizen, G. (2016). A new species of *Charinus* Simon, 1892 (Arachnida: Amblypygi: Charinidae) from Israel and new records of *C. ioanniticus*. *Eur. J. Taxon.* 1–17.
Mittmann, B. and Wolff, C. (2012). Embryonic development and staging of the cobweb spider *Parasteatoda tepidariorum* C. L. Koch, 1841 (syn.: *Achaearanea tepidariorum*; Araneomorphae; Theridiidae). *Dev. Genes. Evol.* 222, 189–216.
Mojaddidi, H., Fernandez, F. E., Erickson, P. A. and Protas, M. E. (2018). Embryonic origin and genetic basis of cave associated phenotypes in the isopod crustacean *Asellus aquaticus*. *Sci. Rep.* 8, 16589–12.
Morehouse, N. I., Buschbeck, E. K., Zurek, D. B., Steck, M. and Porter, M. L. (2017). Molecular evolution of spider vision: new opportunities, familiar players. *Biol. Bull.* 233, 21–38.
Nagata, T., Koyanagi, M., Tsukamoto, H., Saeki, S., Isono, K., Shichida, Y., Tokunaga, F., Kinoshita, M., Arikawa, K. and Terakita, A. (2012). Depth perception from image defocus in a jumping spider. *Science (1979)* 335, 469–471.
Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von and Minh, B. Q. (2015). IQ-TREE: a fast

and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.

Nolan, E. D., López, C. E. S. and Sharma, P. P. (2020). Developmental gene expression as a phylogenetic data class: support for the monophyly of Arachnopulmonata. *Dev. Genes Evol.* 230, 137–153.

Nossa, C. W., Havlak, P., Yue, J.-X., Lv, J., Vincent, K. Y., Brockmann, H. J. and Putnam, N. H. (2014). Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* 3, 9.

Oda, H. and Akiyama-Oda, Y. (2020). The common house spider *Parasteatoda tepidariorum*. *Evodevo* 11, 6.

Paese, C. L. B., Leite, D. J., Schönauer, A., McGregor, A. P. and Russell, S. (2018). Duplication and expression of Sox genes in spiders. *BMC Evol. Biol.* 18, 205–14.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419.

Paulus, H. F. (1979). Eye structure and the monophyly of the Arthropoda. In *Arthropod Phylogeny* (ed. Gupta, A. P.), pp. 299–383. Van Nostrand Reinhold.

Pechmann, M. (2016). Formation of the germ-disc in spider embryos by a condensation-like mechanism. *Front. Zool.* 13, 35.

Pechmann, M., McGregor, A. P., Schwager, E. E., Feitosa, N. M. and Damen, W. G. M. (2009). Dynamic gene expression is required for anterior regionalization in a spider. *Proc. Natl. Acad. Sci. USA* 106, 1468–1472.

Pignoni, F., Hu, B., Zavitz, K. H., Xiao, J., Garrity, P. A. and Zipursky, S. L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development. *Cell* 91, 881–891.

Pineda, D., Gonzalez, J., Callaerts, P., Ikeo, K., Gehring, W. J. and Salo, E. (2000). Searching for the prototypic eye genetic network: *Sine oculis* is essential for eye regeneration in planarians. *Proc. Natl. Acad. Sci. USA* 97, 4525–4529.

Porter, M. L., Dittmar, K. and Pérez-Losada, M. (2007). How long does evolution of the troglomorphic form take? Estimating divergence times in *Astyanax mexicanus*. *Acta Carsologica* 36, 173–182.

Posnien, N., Zeng, V., Schwager, E. E., Pechmann, M., Hilbrant, M., Keefe, J. D., Damen, W. G. M., Prpic, N.-M., McGregor, A. P. and Extavour, C. G. (2014). A comprehensive reference transcriptome resource for the common house spider *Parasteatoda tepidariorum*. *PLoS One* 9, e104885.

Protas, M. E. and Jeffery, W. R. (2012). Evolution and development in cave animals: from fish to crustaceans. *Wiley Interdiscip. Rev. Dev. Biol.* 1, 823–845.

Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R. and Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* 38, 107–111.

Protas, M. E., Trontelj, P. and Patel, N. H. (2011). Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proc. Natl. Acad. Sci. USA* 108, 5702–5707.

Ramirez, M. D., Pairett, A. N., Pankey, M. S., Serb, J. M., Speiser, D. I., Swafford, A. J. and Oakley, T. H. (2016). The last common ancestor of most bilaterian animals possessed at least 9 opsins. *Genome Biol. Evol.* 248 –13.

Re, C., Fišer, Ž., Perez, J., Tacdol, A., Trontelj, P. and Protas, M. E. (2018). Common genetic basis of eye and pigment loss in two distinct cave populations of the isopod crustacean *Asellus aquaticus*. *Integr. Comp. Biol.* 58, 421–430.

Riddle, M. R., Aspiras, A. C., Gaudenz, K., Peuß, R., Sung, J. Y., Martineau, B., Peavey, M., Box, A. C., Tabin, J. A., McGaugh, S., et al. (2018). Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. *Nature* 555, 647–651.

Rivera, A., Winters, I., Rued, A., Ding, S., Posfai, D., Cieniewicz, B., Cameron, K., Gentile, L. and Hill, A. (2013). The evolution and function of the Pax/Six regulatory network in sponges. *Evol. Dev.* 15, 186–196.

Roman, G., He, J. and Davis, R. L. (2000). *kurtz*, a novel nonvisual arrestin, is an essential neural gene in *Drosophila*. *Genetics* 155, 1281–1295.

Rosenberg, M. I., Lynch, J. A. and Desplan, C. (2009). Heads and tails: Evolution of antero-posterior patterning in insects. *Biochim. Biophys. Acta* 1789, 333–342.

Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G. D., Longhorn, S. J., Peterson, K. J., Pisani, D., Philippe, H. and Telford, M. J. (2011). A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc. R. Soc. B* 278, 298–306.

Samadi, L., Schmid, A. and Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *Evodevo* 6, 16.

Sandve, S. R., Rohlfs, R. V and Hvidsten, T. R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat. Genet.* 50, 908–909.

Santos, V. T., Ribeiro, L., Fraga, A., Barros, C. M., Campos, E., Moraes, J., Fontenele, M. R., Araújo, H. M., Feitosa, N. M., Logullo, C., et al. (2013). The embryogenesis of the tick *Rhipicephalus* (*Boophilus*) *microplus*: the establishment of a new chelicerate model system. *Genesis* 51, 803–818.

Schacht, M. I., Schomburg, C. and Bucher, G. (2020). *six3* acts upstream of *foxQ2* in labrum and neural development in the spider *Parasteatoda tepidariorum*. *Dev. Genes Evol.* 230, 95–104.

Schomburg, C., Turetzek, N., Schacht, M. I., Schneider, J., Kirfel, P., Prpic, N.-M. and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *Evodevo* 6, 15.

Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P. and Damen, W. G. M. (2009). *hunchback* functions as a segmentation gene in the spider *Achaearanea tepidariorum*. *Curr. Biol.* 19, 1333–1340.

Schwager, E. E., Meng, Y. and Extavour, C. (2015). *vasa* and *piwi* are required for mitotic integrity in early embryogenesis in the spider *Parasteatoda tepidariorum*. *Dev. Biol.* 402, 276–290.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62.

Setton, E. V. W., March, L. E., Nolan, E. D., Jones, T. E., Cho, H., Wheeler, W. C., Extavour, C. G. and Sharma, P. P. (2017). Expression and function of *spineless* orthologs correlate with distal deutocerebral appendage morphology across Arthropoda. *Dev. Biol.* 430, 224–236.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.

Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. and Extavour, C. G. (2013). *Distal-less* and *dachshund* pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman *Phalangium opilio* (Opiliones). *Evol. Dev.* 15, 228–242.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Wheeler, W. C. (2014a). Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281, 20140661.

Sharma, P. P., Kaluziak, S. T., Pérez-Porro, A. R., González, V. L., Hormiga, G., Wheeler, W. C. and Giribet, G. (2014b). Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31, 2963–2984.

Sharma, P. P., Santiago, M. A., González-Santillán, E., Monod, L. and Wheeler, W. C. (2015a). Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol. Dev.* 17, 347–355.

Sharma, P. P., Tarazona, O. A., Lopez, D. H., Schwager, E. E., Cohn, M. J., Wheeler, W. C. and Extavour, C. G. (2015b). A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. B* 282, 20150698.

Shubin, N., Tabin, C. and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.

Shull, L. C., Sen, R., Menzel, J., Goyama, S., Kurokawa, M. and Artinger, K. B. (2020). The conserved and divergent roles of Prdm3 and Prdm16 in zebrafish and mouse craniofacial development. *Dev. Biol*. 461, 132–144.

Smrž, J., Kováč, Ľ., Mikeš, J. and Lukešová, A. (2013). Microwhip scorpions (Palpigradi) feed on heterotrophic cyanobacteria in Slovak caves - a curiosity among Arachnida. *PLoS One* 8, e75989.

Soneson, C., Love, M. I. and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521.

Stahl, B. A., Gross, J. B., Speiser, D. I., Oakley, T. H., Patel, N. H., Gould, D. B. and Protas, M. E. (2015). A transcriptomic analysis of cave, surface, and hybrid isopod crustaceans of the species *Asellus aquaticus*. *PLoS One* 10, e0140484.

Strickler, A. G., Yamamoto, Y. and Jeffery, W. R. (2001). Early and late changes in *Pax6* expression accompany eye degeneration during cavefish development. *Dev. Genes Evol.* 211, 138–144.

Takagi, A., Kurita, K., Terasawa, T., Nakamura, T., Bando, T., Moriyama, Y., Mito, T., Noji, S. and Ohuchi, H. (2012). Functional analysis of the role of *eyes absent* and *sine oculis* in the developing eye of the cricket *Gryllus bimaculatus*. *Dev. Growth Differ.* 54, 227–240.

Telford, M. J. and Thomas, R. H. (1998). Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* 95, 10671–10675.

Thoen, H. H., How, M. J., Chiou, T.-H. and Marshall, J. (2014). A different form of color vision in mantis shrimp. *Science (1979)* 343, 411–413.

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J. and Prpic, N.-M. (2015). Neofunctionalization of a duplicate *dachshund* gene underlies the evolution of a novel leg segment in arachnids. *Mol. Biol. Evol.* 33, 109–121.

Vopalensky, P. and Kozmik, Z. (2009). Eye evolution: common use and independent recruitment of genetic components. *Phil. Trans. R. Soc. B* 364, 2819–2832.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V and Zdobnov, E. M. (2017). BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548.

Weygoldt, P. (1975). Untersuchungen zur Embryologie und Morphologie der Geißelspinne *Tarantula marginemaculata* C. L. Koch (Arachnida, Amblypygi, Tarantulidae). *Zoomorphologie* 82, 137–199.

Weygoldt, P. (2000). *Whip Spiders (Chelicerata: Amblypygi). Their Biology, Morphology and Systematics.* Apollo Books.

Yang, X., ZarinKamar, N., Bao, R. and Friedrich, M. (2009). Probing the *Drosophila* retinal determination gene network in *Tribolium* (I): The early retinal genes *dachshund, eyes absent* and *sine oculis*. *Dev. Biol.* 333, 202–214.

Young, M. D., Wakefield, M. J., Smyth, G. K. and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14.

ZarinKamar, N., Yang, X., Bao, R., Friedrich, F., Beutel, R. and Friedrich, M. (2011). The *Pax* gene *eyegone* facilitates repression of eye development in *Tribolium*. *Evodevo* 2, 8.

Zhou, Y., Liang, Y., Yan, Q., Zhang, L., Chen, D., Ruan, L., Kong, Y., Shi, H., Chen, M. and Chen, J. (2020). The draft genome of horseshoe crab *Tachypleus tridentatus* reveals its evolutionary scenario and well-developed innate immunity. *BMC Genom.* 21, 137.

Zopf, L. M., Schmid, A., Fredman, D. and Eriksson, B. J. (2013). Spectral sensitivity of the ctenid spider *Cupiennius salei*. *J. Exp. Biol.* 216, 4103–4108.

Zurek, D. B., Cronin, T. W., Taylor, L. A., Byrne, K., Sullivan, M. L. G. and Morehouse, N. I. (2015). Spectral filtering enables trichromatic vision in colorful jumping spiders. *Curr. Biol.* 25, R403–R404.

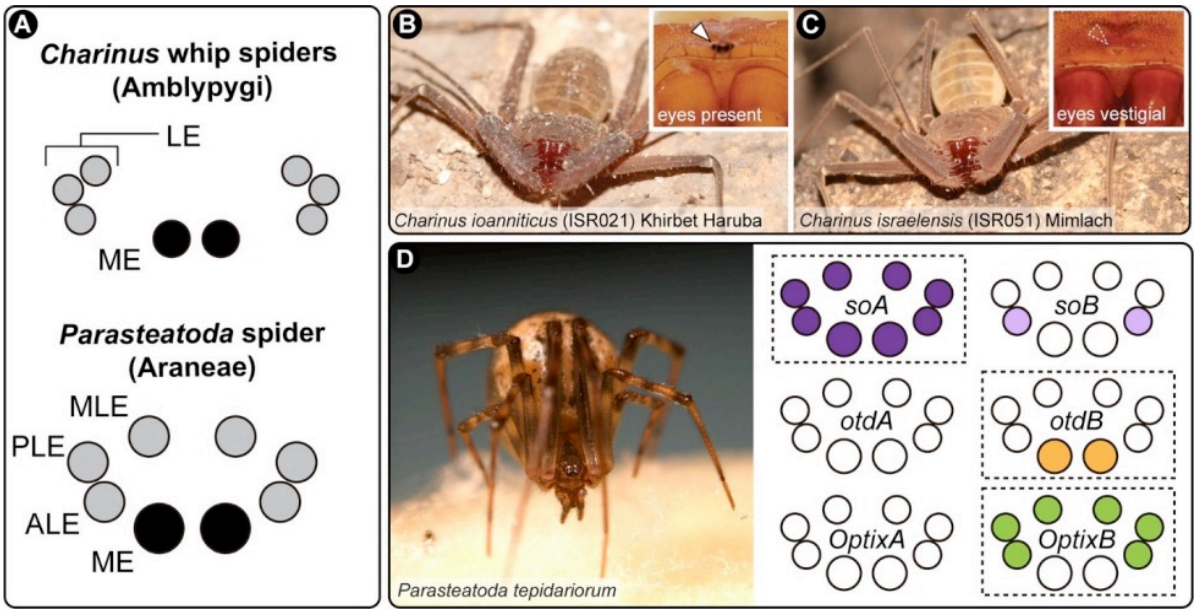


Figure 1. Study species and their corresponding eye arrangements. (a) Schematic representation of the eyes of *Charinus* whip spiders (Amblypygi) (upper), and the spider *Parasteatoda tepidariorum* (Araneae; lower). (b) Live specimen of *C. ioanniticus* from Khirbet Haruba cave (Haruva cave). Inset: detail of the median eyes. (c) Live specimen of *C. israelensis* from Mimlach cave. Inset: detail of the reduced median eyes. (d) Live specimen of *P. tepidariorum*, and schematic representation of the expression patterns of paralog pairs of *Ptep-sine oculis* (*soA*/*soB*), *Ptep-orthodenticle* (*otdA*/ *otdB*), and *Ptep-Optix* (*OptixA*/*OptixB*) in the eyes. ME: median eyes; ALE: anterior lateral eyes; PLE: posterior lateral eyes; MLE: median lateral eyes; LE: lateral eyes. B–C by Shlomi Aharon; D by Jesús A. Ballesteros.

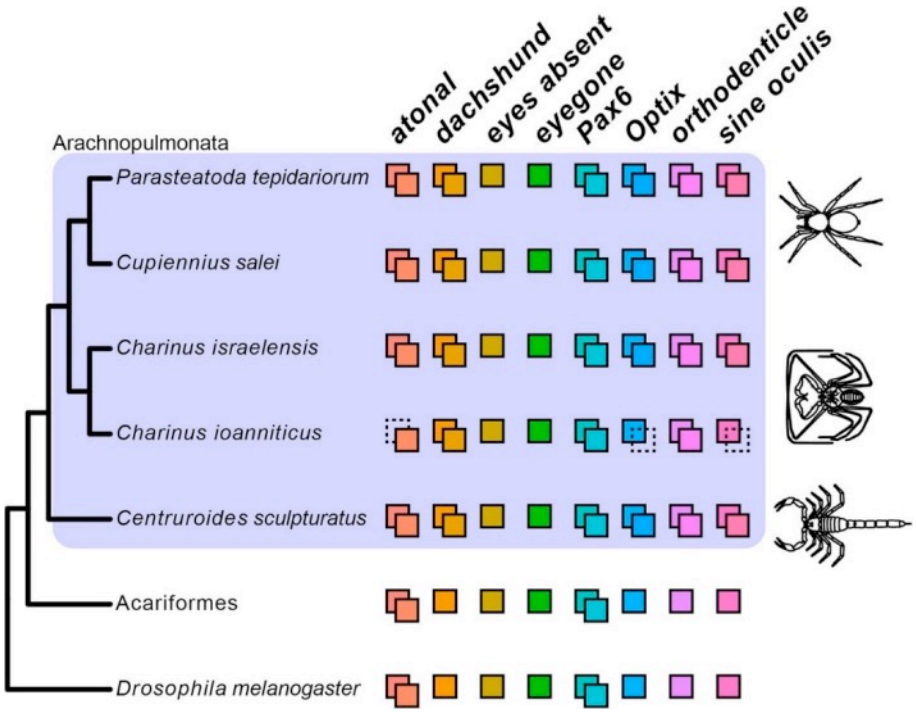


Figure 2. Phylogenetic distribution of Retinal Determination Gene Network (RDGN) homologs in an insect (*Drosophila melanogaster*), a non-arachnopulmonate arachnid group (Acariformes: *Dinothrombium tinctorium; Tetranychus urticae*) and Arachnopulmonata (spider: *Parasteatoda tepidariorum*; scorpion: *Centruroides sculpturatus*), including newly discovered orthologs in *Charinus* whip spiders (Amblypygi). Colored squares indicate number of paralogs for each RDGN gene. Dotted squares indicate presumed missing data, not gene loss. For comprehensive list of duplicated genes in Arachnopulmonata see (Schwager et al., 2017) and (Leite et al., 2018). Gene trees and alignments for each gene are available in Additional file 1 Dataset S1.

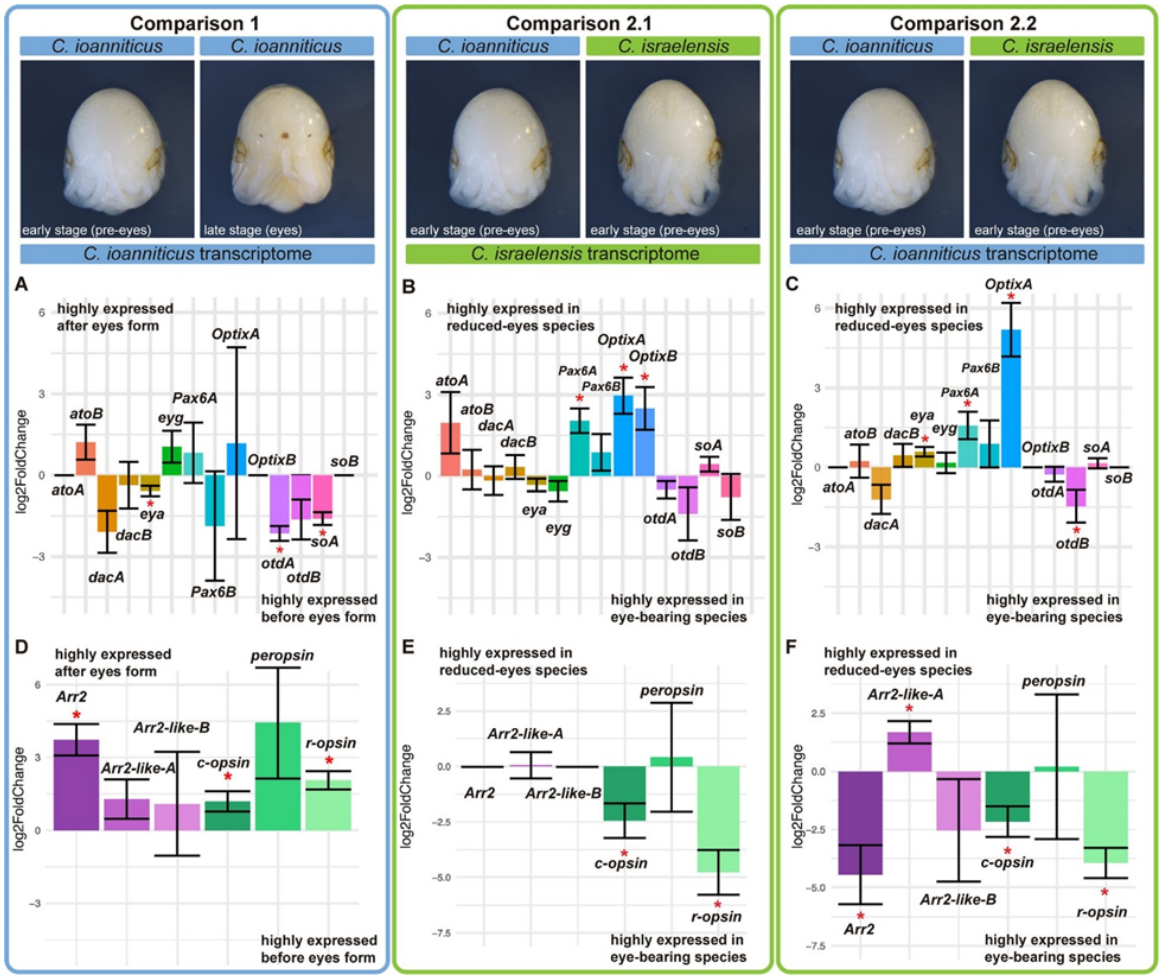


Figure 3. Differential gene expression analysis of Retinal Determination Gene Network (RDGN) genes and phototransduction genes in *Charinus* whip spider deutembryos. Bar graphs display log2 fold change of selected RDGN and phototransduction genes. The denominator of the differential gene expression is the always the sample in the left. (a), (d) Comparison 1; Comparison between reads of early (pre-eyespot) and late deutembryos (eyespot) of the eye-bearing species *C. ioanniticus* mapped onto *C. ioanniticus* transcriptome. (b), (e) Comparison 2.1; Comparison between reads of early deutembryo of *C. ioanniticus* and early deutembryo of *C. israelensis* mapped onto *C. israelensis* transcriptome. (c), (f) Comparison 2.2; Comparison between reads of early deutembryo of *C. ioanniticus* and early deutembryo of *C. israelensis* mapped onto *C. ioanniticus* transcriptome. *AtoA/B: atonalA/atonalB; dacA/B: dachshundA/B; eya: eyes absent; eyg: eyegone; otdA/B: orthodenticleA/B; soA/B: sine oculisA/B. Arr2: Arrestin-2; Arr2- likeA/B: Arrestin-2-like A/B*. Asterisks denote genes that are differentially expressed with a padj > 0.05. Log2FC = 0 for *atoA, OptixB*, and *soB* for Comparison 1 and Comparison 2.2 are due to the absence of those paralogs in *C. ioanniticus* reference transcriptome. Log2FC = 0 for Arr2 in Comparison 2.1 is due to absence of this gene in *C. israelensis* reference transcriptome.

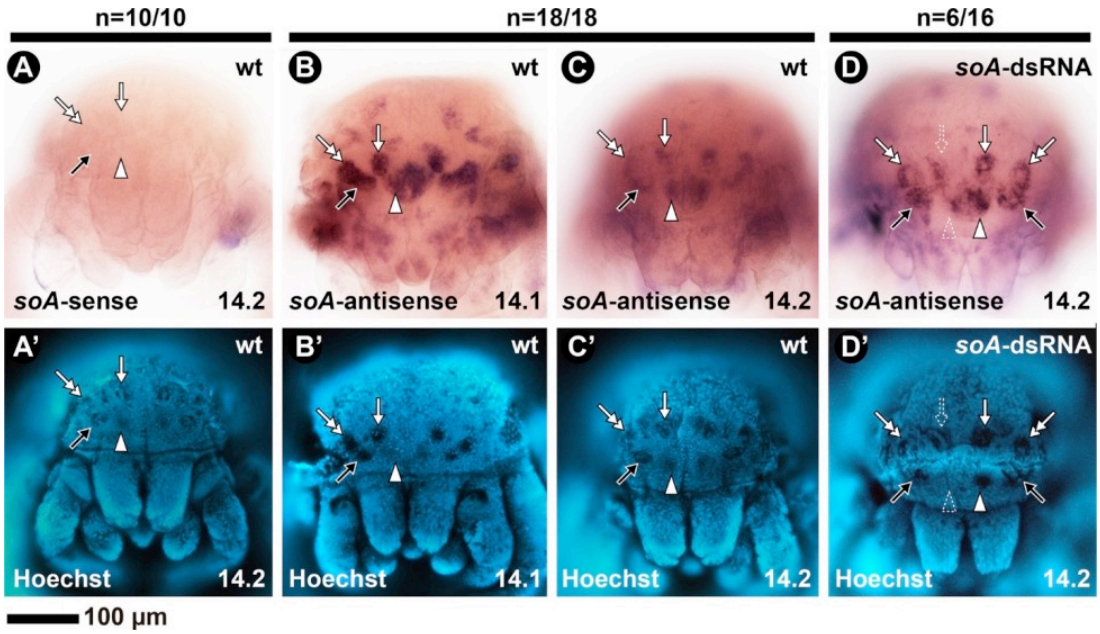


Figure 4. In situ hybridization using DIG-labeled riboprobes for *Ptep-soA* in late embryos of the spider *Parasteatoda tepidariorum*. All embryos in frontal view. (a–d) bright field images. A'–D': Same embryos, in Hoechst staining. (a) Sense probe of a stage 14.2 embryo (no signal).(b) Antisense probe on a wild type stage 14.1 embryo.(c) Antisense probe on a wild type stage 14.2 embryo.(d) Antisense probe on a stage 14.2 embryo from the *Ptep-soA* dsRNA-injected treatment. *soA: sine oculis A*. White arrowhead: median eye; Black arrow: anterior lateral eye; White arrow: median lateral eye; Double white arrow: Posterior lateral eye. Dotted arrowhead/arrow indicate asymmetrical expression and eye defect. Sample sizes are indicated above each treatment.

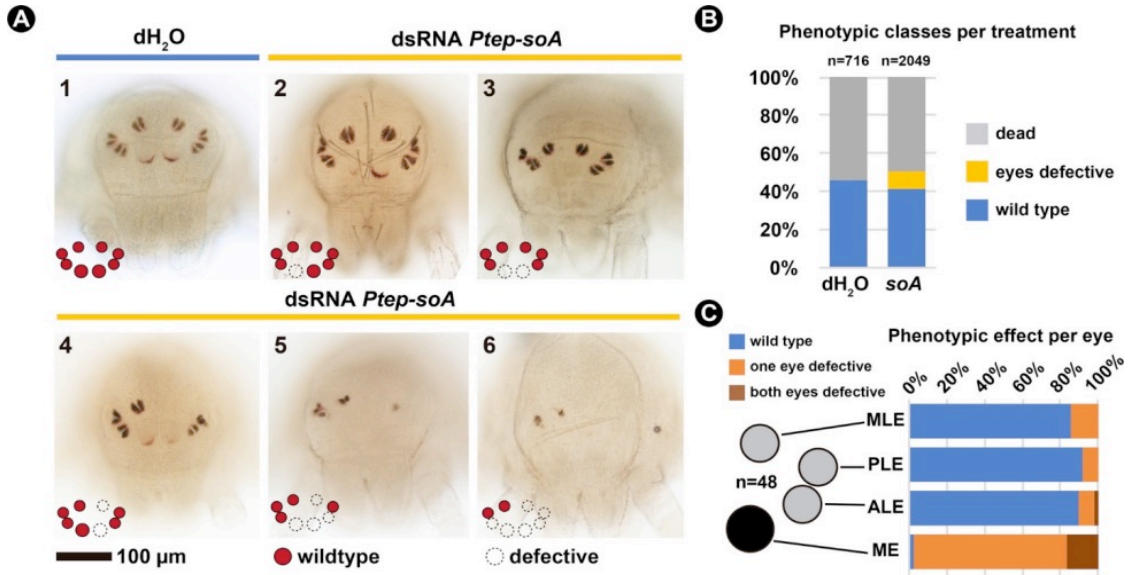


Figure 5. RNA interference against *Ptep-sine oculis A*. (a) Bright field images of the spider *Parasteatoda tepidariorum* postembryos resulting from control treatment (dH2O-injected, panel 1) and double stranded RNA (dsRNA) injected treatment (panels 2–6), in frontal view. (b) Frequencies of each phenotypic class per treatment from the combined clutches of all females. See Additional file 1 Fig. S15 for counts per clutch. (c) Frequencies of symmetrical, asymmetrical, and wild type eyes quantified from a subset of 48 individuals with eye reduction phenotype. See Additional file 1 Fig. S14 for figures of all specimens and coding, and Methods for the scoring criteria. ME: median eyes; ALE: anterior lateral eyes; PLE: posterior lateral eyes; MLE: median lateral eyes. Schematics for the different eye types follows the nomenclature in Fig. 1.

Chapter 6

Vestigial organs alter fossil placements in an ancient group of terrestrial chelicerates

Guilherme Gainett[1]; Benjamin C.[1] Klementz, Pola Blaszczyk[1], Emily V. W. Setton[1], Gabriel P. Murayama[3], Rodrigo Willemart[3], Efrat Gavish-Regev[2]; Prashant P. Sharma[1]

[1]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, 53706, USA.
[2]The National Natural History Collections, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 9190401, Israel.
[3]Laboratório de Ecologia Sensorial e Comportamento de Artrópodes, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Rua Arlindo Béttio, 1000, Ermelino Matarazzo, São Paulo, SP 03828-000, Brazil.

Unpublished Chapter

Summary paragraph

The importance of paleontological data for inferring phylogenetic relationships of living taxa has been debated extensively (Cobbett et al., 2007; Donoghue et al., 1989; Patterson, 1981). An accruing body of empirical works has shown that adding fossil data can deeply alter tree topologies (Coiro et al., 2018; Davesne et al., 2016; Donoghue et al., 1989), often more so than morphological data of extant taxa (Asher et al., 2019; Beck and Baillie, 2018; Koch and Parry, 2020). The influence of latent characters of extant taxa, such as vestigial and rudimentary organs, on phylogenetic placement is poorly understood, as vestigial organs are often overlooked or treated as absent. Here we show that vestigial lateral eyes occur in extant harvestmen, a group once thought to possess only one pair of median eyes. Neuroanatomical gene expression surveys of eye-patterning transcription factors, opsins, and structural proteins in the daddy-longlegs *Phalangium opilio* show that the vestigial lateral eyes are innervated to a region of the brain that is positionally homologous to the lateral eye neuropil of arachnids. Gene silencing of *eyes absent* shows that the vestigial lateral eyes are under the control of the retinal determination network. The incidence of lateral eyes in extant species

bears upon the placement of the oldest harvestmen fossils, stem-groups that possess both a pair of median eyes and a pair of lateral eyes. Sampling embryos of a tropical harvestman species, we show that vestigial lateral eyes were likely present in the common ancestor of all lineages thought to bear only median eyes. Phylogenetic analysis of harvestman relationships with updated understanding of lateral eye incidence resolved the four-eyed fossil group as a member of an extant daddy-longlegs suborder, which in turn resulted in older estimated ages of harvestman diversification under total evidence dating. This work reveals that vestigial structures of extant taxa have the potential to influence character polarity, placement of fossils, and inference of divergence times.

Main

Rudimentary organs have long fascinated evolutionary biologists for their potential to inform common ancestry and macroevolutionary processes (Cohn and Tickle, 1999; Darwin, 1859; Hall, 2003). Arthropod eyes exhibit a complex evolutionary history, with losses and gains occurring repeatedly (Henze and Oakley, 2015). Chelicerates (sea spiders, horseshoe crabs and other arachnids) have two types of visual systems, the median eyes (principal eyes) and the lateral eyes (secondary eyes), which differ in their mode of embryogenesis, acuity, and connection to the brain (Battelle, 2006; Land, 1985). Evidence from position, genetic architecture, and morphology suggests that the median eyes of chelicerates are homologous to the ocelli of insects, and the lateral eyes are the homologs of the facetted (compound) eyes (Friedrich, 2006; Morehouse et al., 2017; Paulus, 1979). Within Chelicerata, there is considerable diversity among orders concerning the number of eyes and reduction of visual systems, with only horseshoe crabs and some fossil arachnids exhibiting facetted eyes (either as compound or semi-compound eyes) (Miether and Dunlop, 2016).

Extant harvestmen (Opiliones) have a maximum of one pair of single-lens camera-type eyes (Shultz and Pinto-da-Rocha, 2007). In Phalangida (daddy-longlegs and allies), these occur as median eyes placed on a median body protrusion, the ocularium (Fig. 1a) (Lehmann et al., 2016; Shultz and Pinto-da-Rocha, 2007). In Cyphophthalmi (mite harvestmen), species of most families are blind, but some genera of Styllocellidae and Pettalidae retain a pair of laterally placed eyes associated with a lateral protrusion of the body, the ozophore (Fig. 1a) (Sharma and Giribet, 2006; Alberti et al., 2008). Recently, a putative stem group fossil was discovered that exhibits two pairs of eyes, with each pair associated with an ocularium and an ozophore, suggesting that the common ancestor of Opiliones possessed both visual systems (Fig. 1a) (Garwood et al., 2014). The ensuing evolutionary scenario postulated that all Cyphophthalmi have lost median eyes, whereas all Phalangida have lost lateral eyes.

We aimed to investigate the molecular mechanisms underpinning the loss of lateral eyes in Phalangida, focusing on the tractable model *Phalangium opilio*. Unexpectedly, we discovered rudimentary eyes in this species, which led us to reassess the phylogenetic position of the stem group four-eyed harvestman and the attendant scenario of eye reduction in Opiliones.

An extant daddy-longlegs has multiple rudimentary eye pairs

The retinal determination network (RDN) includes important genes conserved across bilaterians in initiating and patterning photoreceptive structures, such *Pax6*, *eyes absent* (*eya*), and *sine oculis* (*so*) (Gehring, 2001; Kumar, 2009). In spiders, most RDN genes are expressed in subdomains of the developing eyes (Samadi et al., 2015; Schomburg et al., 2015). We selected seven genes important for eye morphogenesis and photoreceptor specification in the ocelli and compound eyes of the fruit fly *Drosophila melanogaster* (Kumar, 2009), and whose

homologs exhibit expression in the median and lateral eyes of other arachnids, such as spiders (Fig. 1b; Extended Data Fig. 1f–m'; Extended Data Fig. 2–4) (Samadi et al., 2015; Schomburg et al., 2015). A detailed description of the expression patterns of the seven genes is provided in the Supplementary Results.

The median eyes of the harvestman *P. opilio* develop as two large bilateral eye folds which grow anteriorly over the neurogenetic ectoderm (Fig. 1b; Extended Data Fig. 1a–e). When fully folded, the eye field is a multilayered tissue above the developing brain (Moritz, 1957) (Extended Data Fig. 1f–m'; Supplementary Results). Ubiquitous spatial and temporal expression of *Po-eya* and *Po-so* (Fig. 1b; Extended Data Fig. 2–4), as well as of *Po-otd* (Extended Data Fig. 4), was associated with all phases of median eye development, suggesting a pivotal role for these genes in the establishment and differentiation of the median eyes of *P. opilio*.

Beyond the median eyes, expression of *Po-eya* occurs within a broader domain of *Po-so* expression on the anterolateral margin of head, ectal to the lateral furrow (Fig. 1b; Extended Data Fig. 2). This region corresponds to the sites of lateral optical neuropil and lateral eye development in spiders (Doeffinger et al., 2010; Samadi et al., 2015; Schomburg et al., 2015). These *Po-eya* domains temporally precede foci of *Po-otd* expression, which is co-expressed with *Po-eya* in these cells during the stages investigated (Extended Data Fig. 4). *Po-eya* and *Po-so* expression also outline the developing lateral region of the brain (hereafter, lateral brain center) (Fig. 1b; Extended Data Fig. 2).

The co-expression of the three main eye patterning genes where lateral eyes normally form in arachnids prompted us to investigate whether *P. opilio* develops rudimentary eyes. We investigated the expression of opsin-encoding genes, as opsins are conserved proteins upstream of the photoreceptive cascade in photoreceptive organs across Metazoa (Ramirez et al., 2016). Long wave-length sensitive (LWS) r-opsins are known to be canonical visual opsins in

arachnids (Nagata et al., 2012) and peropsin has previously been detected in the retina of spiders (Nagata et al., 2009). As a point of comparison, we first visualized embryonic expression of LWS r-opsins and a peropsin in two arachinid groups possessing both median and lateral eyes, namely a spider (*Parasteatoda tepidariorum*) and a scorpion (*Centruroides sculpturatus*). In the spider *P. tepidariorum*, peropsin (*Pt-per*) is expressed in the median and lateral eyes of late embryos (stage 14), but LWS r-opsin mRNA (*Pt-Rh1*; *Pt-Rh2*) is not embryonically expressed in the eyes (Fig. 1c). In the scorpion *C. sculpturatus*, peropsin (*Cs-per*) was not detected in the embryonic eyes, whereas one LWS r-opsin paralog (*Cs-Rh1*) was expressed mostly in the median eyes, and a second LWS r-opsin paralog (*Cs-Rh2*) is expressed only in the lateral eyes (Fig. 1d). These results show that embryonic opsin expression serves as markers specific for both median and lateral eyes, despite some lability of opsin use.

Eight opsin genes occur in the genome of *P. opilio* (Extended Data Fig. 5). We identified transcripts predicted as three c-opsins (non-visual), a peropsin (*Po-per*), and four r-opsins (*Po-Rh1*, *Po-Rh3*, *Po-Rh7*, and *Po-arthropsin*) (Extended Data Fig. 5). In embryonic stages shortly prior to eye pigment formation (stages 13–14), we detected *Po-per* expression in the median eyes, and *Po-Rh1* (a LWS r-opsin) in a group of cells in the lateral margin of the head (Fig. 1e), co-localizing with *Po-eya* in the lateral cells (Extended Data Fig. 6a). This segregation of *Po-per* and *Po-Rh1* expression between the median eyes and the lateral cells persists until stage 18, when *r-opsin* also becomes expressed in the median eyes (Extended Fig. 6b). Unexpectedly, *Po-Rh1* is additionally expressed in two small groups of cells anterior to the median eyes (hereafter median cells), which also do not express *Po-per* initially (Fig. 1e; Extended Data Fig. 6b–i). *Po-Rh3* (UV opsin) is expressed in the developing median eyes, lateral cells and median cells, whereas we did not detect expression of *Po-Rh7* opsin and *Po-arthropsin* throughout embryogenesis (Extended Data Fig. 6e). The opsin-positive lateral cells are also present in post-embryonic stages (Extended Data Fig. 6j–k).

Visual arrestins (beta-arrestins) are important downstream components of the photoreception cascade (Alvarez, 2008) and myosin-III (*myoIII*) (homolog of *D. melanogaster ninaC*) is a photoreceptor-specific marker in the chelicerate horseshoe crab known to localize to the developing larval and adult lateral and median eyes (Battelle et al., 2001; Harzsch et al., 2006). In the genome and transcriptomes of *P. opilio*, we discovered two putative visual arrestins (*Po-arrestin-2* and *Po-arrestin-2-like*) and two myosin-III homologs (*Po-myoIII-1*; *Po-myoIII-2*). All markers (except *Po-arrestin-2-like*; no expression detected) are expressed in the median eyes and median cells, whereas *Po-arrestin-2* and *Po-myoIII-2* are also expressed in the lateral cells (Extended Data Fig. 6c–d, 6f–h). The combined expression of a LWS r-opsin, a UV opsin, and downstream photoreceptor-specific genes strongly suggests that the lateral and central cells are photoreceptive. Combined with the expression data of RDN genes, these data support that the lateral and median cells are rudimentary eyes.

Among chelicerates, the basally branching Pygnogonida (sea spiders) have two pairs of median eyes and no lateral eyes, whereas euchelicerates (arachnids and horseshoe crabs) mostly possess both sets of visual systems, with secondary reduction of one or both eye types in miniaturized taxa. Horseshoe crabs are unique in that they possess additional sets of rudimentary (or larval) eyes: one pair of *rudimentary lateral eyes*, which becomes accessory to the facetted eyes in adults; and one pair of *rudimentary median eyes*, which post-embryonically fuse into a subcuticular single eye accessory to the pair of median eyes of adults. Although the number of lateral eye pairs is variable in extant arachnids (0–5 pairs), all extant arachnids are thought to have a maximum of one pair of median eyes. Given the position of the lateral and median cells, we hypothesized that these photoreceptors of *P. opilio* constitute homologs of the rudimentary lateral eyes (i.e., larval lateral eyes of horseshoe crabs) and rudimentary median eyes (i.e., larval median eyes of horseshoe crabs), respectively.

Median and rudimentary eyes of P. opilio *are under shared control of* eyes absent

To test this hypothesis of eye homology, we investigated whether genes patterning all eyes (median and lateral) in other arthropods are necessary for the development of the discovered photoreceptors in *P. opilio*. In *Drosophila melanogaster*, *eya* and *so* are necessary for compound eye and ocelli development (Bonini et al., 1993; Cheyette et al., 1994), and form a protein complex that works in synergy (Pignoni et al., 1997). Previous work in spiders suggested that both median and lateral eye formation is under the control of *eya* and *so* (Samadi et al., 2015; Schomburg et al., 2015), and knockdown of paralog *so-A* in the spider *P. tepidariorum* affects the development both eye types (there are no functional data for *eya* in spiders) (Gainett et al., 2020). In *P. opilio*, we focused on *eya* function, given ubiquitous and early *eya* expression in the median eyes and in the early foci of lateral expression that prefigures the location of *Po-Rh1* expression in the lateral cells.

We conducted RNA interference (RNAi) against *Po-eya* via embryonic injections of double-stranded RNA (dsRNA). Hatchlings in the control treatment (injected with vector dsRNA or water) had wildtype eyes (100% of hatchlings) (Fig. 2a). In the *Po-eya* dsRNA-injected treatment, 83% of hatchlings presented medium eye defects, in a spectrum from smaller median eyes with reduced pigmentation, to absence of the eye and pigmentation (Fig. 2b–c; Extended Data Fig. 7). In embryos, the eye fields were smaller than in controls, and showed correlated, diminished expression of *Po-eya* (Fig. 2d–f). In addition to defects in the median eyes, embryos scored as *eya* phenotypes before in situ hybridization also showed reduced or absent lateral and median cells, as marked by *Po-Rh1* expression (n=10/10 embryos). Notably, the degree of reduction of median eye, median cells and lateral cells was tightly correlated, as evidenced by mosaic embryos (Fig. 2e). These results underscore the

requirement of *eya* for the patterning of the median eye and both sets of rudimentary eyes, and supports their homology with lateral and median eye complexes, respectively.

P. opilio *rudimentary eyes retain the protocerebral innervation pattern of separate visual systems*

The median and lateral eyes of arthropods are innervated by distinct regions of the protocerebrum, the anterior-most component of the tripartite brain. In extant chelicerates with both median and lateral visual systems, the axons of the median eyes connect to rostral neuropils of the protocerebrum, whereas as the lateral eyes connect to separate neuropils that arise from lateral domains of the protocerebrum (Battelle, 2006; Brenneis, 2022; Doeffinger et al., 2010; Lehmann and Melzer, 2013; Lehmann et al., 2016). This innervation pattern is conserved in Arthropoda since the Cambrian, as evidenced by the preserved neuroanatomy of stem euarthropod leanchoillid fossils (Lan et al., 2021).

To establish whether the rudimentary eyes indeed exhibit phylotypic patterns of innervation, we examined the developing brain of *P. opilio*. The embryonic protocerebrum of *P. opilio* (stage 15) is composed of three main regions: a rostral median center, a lateral center (ventrolateral after stage 16) that develops from the lateral furrow, and the posterior arcuate body. These three regions are recognizable by combinations of neural genes (e.g., *Po-otd*; *Po-Pax6b*; Supplementary video S1) and proteins enriched in neural cells (acetylated-tubulin; HRP) (Fig. 3a; Extended Data Fig. 8a–g). Cellular projections extending from the retinal cells of the median eyes form a nerve that projects dorsally toward the median center of the brain (Extended Data Fig. 8e–g; Supplementary video S2), in a position comparable to what has been described using cobalt and DiI/DiO retro-labelling in the eyes of adult daddy-longlegs (Lehmann et al., 2016). Expression of *Po-Rh1* and *Po-Rh3* opsins reveals that fine cellular

projections exist between the medial cells and the median eye (Fig. 3b; Extended Data Fig. 6i). Expression of photoreception-specific markers also reveal that the lateral cells project toward the lateral region of the head, adjacent to the median eyes (Fig. 3c; Extended Data Fig. 6b, e). This projection terminates in the lateral center of the brain (Extended Data Fig. 1b–c), as visualized by double-labeling of *Po-Rh1* and acetylated tubulin antibody, which is enriched in neural cells (Fig. 1d–f; Supplementary video S3). These results show that the rudimentary eyes are connected to the protocerebrum in a pattern consistent with the rudimentary median and lateral eye hypothesis.

Hastocularis argus, a four-eyed fossil, is a crown-group daddy-longlegs

The phylogeny of extant harvestmen is currently resolved into four suborders: Cyphophthalmi (mite harvestmen), Laniatores (armored harvestmen), Eupnoi (daddy-longlegs), and Dyspnoi, with the latter three groups comprising Phalangida (Fernández et al., 2017) (Fig. 1a). The fossil *Hastocularis argus* presents characteristics of both Cyphophthalmi and Phalangida. The presence of an ozophore, the lateral eye, and the open gonostome align it with Cyphophthalmi; the ocularium, the median eye, the presence of a penis, and the elongated legs with tarsomeres (subdivisions of the distal podomere, the tarsus) are clear features of Phalangida (Shultz and Pinto-da-Rocha, 2007). Previously, *H. argus* and the oldest known Rhynie Chert Opiliones fossil, *Eophalangium sheari*, were recovered in a clade as stem-group Cyphophthalmi, based on total evidence analysis (Garwood et al., 2014; Sharma and Giribet, 2014). This placement suggested that intromittent genitalia and elongate legs were plesiomorphic in Opiliones, with secondarily reversals in Cyphophthalmi to regain archetypal arachnid plesiomorphies (e.g., spermatophores; robust legs with undivided tarsi), which is a counterintuitive and non-parsimonious scenario.

A notable limitation of total evidence analysis with small-bodied arthropod fossils is the paucity of observable morphological data, with ensuing placements highly contingent upon a small number of characters. We suspected that the previous placement of *H. argus* and *E. sheari* as stem-group Cyphophthalmi hinged upon the assumption that true lateral eyes occur only in Cyphophthalmi, whereas true median eyes occur only in Phalangida. The discovery of rudimentary lateral eyes in the daddy-longlegs *P. opilio* challenges this paradigm of absence lateral eyes in Phalangida and prompted us to reinvestigate this character and the phylogenetic placement of *H. argus*.

To test whether the rudimentary eyes we found in *P. opilio* occur more broadly across Phalangida, we collected embryos, generated genomic resources, established expression protocols, and assayed opsin expression in the armored harvestman *Iporagaia pustulosa* (a member of Laniatores, the sister group to the remaining Phalangida). Our assays unveiled that rudimentary lateral eyes occur during embryogenesis of this species as well, as marked by the expression of LWS Rh1 opsin (Fig. 4c–d). We found no evidence of median cells in this species. Our results suggest that rudimentary lateral eyes were present in the common ancestor of daddy-longlegs and armored harvestmen, and by extension, likely occur across extant Phalangida.

Next, we augmented the total evidence dataset of (Sharma and Giribet, 2014) by adding *I. pustulosa* and enriching the molecular sequence matrix to reduce missing data in extant terminals. In addition to implementing the original coding (only median eyes present in Phalangida), we recoded presence of lateral eyes in Phalangida under two schemes: (1) under strict coding, lateral eyes were scored as present in *P. opilio* (Eupnoi) and *I. pustulosa* (Laniatores), and all other Phalangida were scored as missing data for this trait; (2) under ground plan coding, lateral eyes were scored as present in all Phalangida, excepting troglobitic taxa (scored as missing data for lateral eyes). Whereas the original coding scheme recovered

the previous relationship of *H. argus* and *E. sheari* as stem-group Cyphophthalmi (Extended Data Fig. 9a), both the strict and the ground plan coding schemes resolved *H. argus* and *E. sheari* as nested within the Eupnoi (Fig. 4a–b; Extended Data Fig. 9b). These placements are consistent with a single origin of intromittent genitalia and leg elongation in the common ancestor of Phalangida.

To assess the impact of these new placements on divergence time estimation under a Bayesian framework, we performed a total evidence molecular dating using the original coding scheme and the strict coding scheme. Under the updated morphological matrix, the median crown group of age of Eupnoi (daddy-longlegs) increased by 88 Myr compared to the traditional coding (Extended Data Fig. 10 a–b). The inferred ages for Phalangida and Opiliones increased by 19 Myr and 14 Myr, respectively (Extended Data Fig. 10a–b). These results suggest that harvestmen are older than previously inferred and that crown-group Phalangida (e.g., *E. sheari*) were an established part of terrestrial Devonian ecosystems.

Vestigial eyes inform evolution of visual systems across Arthropoda

The homology of median eyes across Arthropoda, and especially so the ocelli of Hexapoda and the frontal eyes of crustaceans, is disputed, given variation in the number of eyes and the innervation of the protocerebrum (Elofsson, 2006; Paulus, 1979; Strausfeld et al., 2016). In crustaceans, there are three conserved naupliar (larval) eyes, and up to four additional frontal eyes (“frontal organs”) (Elofsson, 2006). Insects typically have three dorsal ocelli (Baird, 2023; Paulus, 1979), but Collembola may have additional frontal photoreceptors (Paulus, 1979). On the other hand, the conserved visual pathway of the median eyes in sea spiders (Pycnogonida), horseshoe crabs (Xiphosura) and other arachnids strongly supports the homology of all median eyes across Chelicerata (Battelle, 2006; Brenneis, 2022; Lehmann and

Melzer, 2013). Nonetheless, the ancestral number of median eyes in Chelicerata is unclear, as terrestrial arachnids have a maximum of two median eyes. This inference of ancestral state is further confounded by the ongoing debate over the phylogenetic position of horseshoe crabs, as either sister group to the rest of the arachnids, or as nested withing Arachnida (Ballesteros et al., 2022; Lozano-Fernandez et al., 2019). Regardless of the phylogenetic position of horseshoe crabs, the discovery of an additional pair of rudimentary median eyes in the ontogeny of daddy-longlegs, together with the condition of two median eye pairs in Pycnogonida, implies that four median eyes is part of the chelicerate ground plan.

Under this hypothesis, embryos of other extant arachnid orders, as well as stem group euchelicerate fossils, are inferred to exhibit two median eye pairs. Of special interest are the Cambrian putative stem-group chelicerates, such *Mollisonia* and sanctacariids (Legg, 2014; Ortega-Hernández et al., 2022). Paralleling the discovery of rudimentary median eyes in *P. opilio* in this study, trilobites have recently been reported to bear three median eyes, a state comparable to many of the extant mandibulates (e.g., hexapods and crustaceans). The median eyes of trilobites are thought to have been overlooked because of their occurrence in immature stages and their subcuticular nature (Schoenemann and Clarkson, 2023). A better understanding of median eye evolution in Chelicerata may therefore require both novel paleontological and embryological data.

Fossils are increasingly understood to be decisive for phylogenetic reconstruction. Our study suggests that exploration and incorporation of rudimentary organ data in integrative phylogenies can contribute decisively toward character polarity in deep nodes in the tree of life, and thereby resolve the position of contentious fossils.

Methods

Embryo collection and fixation

Adult *Phalangium opilio* were collected between 2017 and 2022 in Madison, Wisconsin (USA). Animal rearing, collection and fixation follow protocols detailed previously (Gainett et al., 2022; Sharma et al., 2012a). Embryos of the laniatorean harvestmen *Iporangaia pustulosa*, *Neosadocus maximus* and *Ampheres leucopheus* (Gonyleptidae) were collected on January 2023 from leaves at State Park Intervales, Ribeirão Grande-SP (Brazil), and fixed as described for *P. opilio*. For *I. pustulosa* transcriptome assembly, a range of embryonic stages was fixed in RNA*later* in the field and transferred to Trizol for total RNA extraction (Thermo Fischer). Library preparation (TrueSeq stranded mRNA) and sequencing was performed by the Biotechnology Center of UW-Madison (USA) in an Illumina NovaSeq 6000 with a 150bp paired-read strategy. Assembly was performed with Trinity v. 2.15.1 (Grabherr et al., 2011). Females of the scorpion *Centruroides sculpturatus* were hand-collected from sites in Arizona (USA) by citizen-scientist collaborators. Embryos were dissed from gravid females in 1x PBS, fixed for 20 minutes in 4% formaldehyde in 1x PBS, washed in 1x PBST and gradually dehydrated into 96% ethanol. Embryos of the spider *Parastetoda tepidariorum* were obtained from a laboratory colony in Madison-Wisconsin (USA) and fixed as described above for *P. opilio*.

Gene isolation

P. opilio orthologs of RDN genes *sine oculis* (*Po-so*), *Pax6a* (*Po-Pax6a*), *orthodenticle* (*Po-otd*), and *dachshund* (*Po-dac*) were previously identified from embryonic transcriptomes

(Garwood et al., 2014; Sharma et al., 2012b). Orthologs of *eyes absent* (*Po-eya*), *Pax6b* (*Po-Pax6b*), and *Optix* (*Po-Optix*), were identified from the *P. opilio* genome annotation (Gainett et al., 2021). All seven genes were additionally identified from either the available embryonic transcriptomes or genomic scaffolds using tblastn (Altschul et al., 1990), to cross validate their sequences and obtain the most complete transcript for downstream applications (Supplementary Table 1). Sequences were reciprocally blasted against NCBI protein database.

Candidate opsins were identified from the *P. opilio* genome annotation and two embryonic transcriptomes using tblastn with a protein query of the spider *P. tepidariorum* opsin sequences (seven opsins) identified by a previous study (Schomburg et al., 2015) (Supplementary Table 1). Candidate opsins were identified from the transcriptomes of the harvestmen *A. leucopheus*, *I. pustulosa*, *N. maximus*, and *A. acuta*, and genome annotation of the scorpion *C. sculpturatus*, via tblastn using opsin protein sequences of *P. opilio* (8 opsins) and *P. tepidariorum* as queries (Supplementary Table 2). All blast results were retrieved and subject to protein predictions with TransDecoder v. 5.0.1 (Bryant et al., 2017), using a minimum length of 50 amino acids (-m 50). All proteins were aligned to a small opsin dataset from Schomburg et al. (Schomburg et al., 2015) with Clustal Omega (Sievers et al., 2011) and a preliminary maximum likelihood analysis was performed with FastTree v. 2.1.10 (Price et al., 2010) to exclude non-opsin sequences. Genes nested in the opsin clades were selected and aligned as described above to a large dataset of metazoan opsins from previous studies (Battelle et al., 2016; Morehouse et al., 2017) and analysed under maximum likelihood with IQ-TREE (Nguyen et al., 2015) (IQ-TREE v.1.6.10, *–mset LG+F+G4 –bb 1000*). Nomenclature of arachnid opsins follows (Morehouse et al., 2017).

Phalangium opilio Po-arrestin and *Po-MyoIII* orthologs were identified from the *P. opilio* genome annotation via tblastn using the horseshoe crab *Limulus polyphemus* visual arrestin (NCBI NP_001301010.1) and myosin III (NCBI AAC16332.3) proteins as query

(Battelle et al., 2001; Smith et al., 1995). We identified one *P. opilio* ortholog of the horseshoe crab visual arrestin (*Po-arrestin-2*; ortholog of *Drosophila melanogaster Arr2* (Gainett et al., 2020)), and two myosin III paralogs (*Po-myoIII-1*; *Po-myoIII-2*) (Supplementary Table 1).

RNA interference (RNAi) via double-stranded RNA (dsRNA) embryonic injections

A 921 bp fragment of *Po-eya* spanning part of the coding sequence was amplified from a cDNA library using gene-specific primers designed with Primer3 v. 4.1.0 (Koressaar and Remm, 2007) and appended with T7 ends. This fragment was linked to a plasmid vector into competent *Escherichia coli* with the TOPO TA Cloning Kit One Shot Top 10 (Invitrogen, CA, USA) following the standard protocol. Plasmids were purified with a GeneJET kit (Thermo Fisher, MA, USA) and sanger sequenced for verification (Supplementary Table 3).

dsRNA was synthetized with the MEGAScript T7 kit (Thermo Fisher, MA, USA). Two clutches of embryos (n=608 embryos) were used for *Po-eya* dsRNA embryonic (knockdown; n=448) and vector dsRNA (control; n=160) embryonic injections (Extended Data Fig. 7). Injection volume was prepared with Rhodamine dextran (1:20) for visualization, at a final concentration between 4–4.5 μg/μL. The experiment was replicated in five additional clutches (*Po-eya* dsRNA; dH_20 control), which were fixed and exclusively used in fluorescent *in situ* hybridization. Clutches of embryos were injected between stage 5 (germ disc) and stage 8, which is before the beginning of the head ectoderm folding process that forms the brain and the eyes. Hatchlings from treatment and control experiments were sorted into four categories of phenotypes: wild type; both eyes lost; one eye small, other eye lost; both eyes small; one eye small, other eye wild type. Only embryos that hatched or nearly hatched (stages 18, 19) were considered in the counts.

Hybridization Chain Reaction (HCR) in situ hybridization and immunochemistry

Hybridization Chain Reaction (HCR) *in situ* hybridization followed a modified version of the Molecular Instruments (Los Angeles, CA, USA) hybridization chain reaction (HCR) v.3 protocol (Bruce et al., 2021; Choi et al., 2018). Probe sequences were designed and synthesized by Molecular Instruments or in an open-source software (Kuehn et al., 2022) and ordered from IDT (USA). Probe catalog numbers, sequences and details about initiators are available in the Supplementary Data 1. Combined HCR and immunochemistry staining followed the protocol in (Bruce et al., 2021; Gainett et al., 2022). The following antibodies were used: (1) primary antibody: goat anti-HRP (123-605-021; Jackson ImmunoResearch Laboratories); (2) primary antibody: mouse acetylated α-tubulin (T6793; Sigma-Aldrich) at 1:500 dilution; (3) secondary antibody: goat anti-mouse Alexa Fluor 488-conjugated (115-545-003; Jackson ImmunoResearch Laboratories) at 1:200 dilution. Imaging was performed on a Zeiss 710 and Zeiss 780 confocal microscope at the Newcomb Imaging Center, UW-Madison, USA. Unless otherwise noted, fluorescent microscopy images are maximum intensity projections, which were linearly adjusted for brightness and contrast in FIJI (v. 2.9.0/1.53t) avoiding overexposure of pixels. Brightfield images were obtained on a Nikon SMZ25 fluorescent stereomicroscope with a DS-Fi2 digital color camera driven by Nikon Elements software. Plates were assembled in Adobe Illustrator 2022. All raw files of confocal stacks (.czi) are available in FigShare.

Phylogenetic analyses and total evidence dating

Total evidence phylogenetic analysis was performed by augmenting a previously established data matrix (Sharma and Giribet, 2014) consisting of 158 morphological characters and five loci (16S rRNA, 28S rRNA, 18S rRNA, cytochrome *c* oxidase subunit I, and histone

H3). To maximize available molecular data, new sequences were added from NCBI (Supplementary Table 4) or retrieved from the embryonic transcriptome of *I. pustulosa* (to be deposited in Genbank) via blastn searches. Multiple sequence alignment and trimming of ambiguously aligned ribosomal regions followed the procedures established by the original matrix with MUSCLE v.3.2 (Edgar, 2004) and GBlocks v.0.91b (Castresana, 2000). In addition, *I. pustulosa* was coded for all 158 morphological characters in the matrix. The final matrix consisted of 52 terminals (6 fossil, 46 extant).

Three alternative morphological coding schemes were implemented. First, reflecting the traditional understanding of harvestman eye evolution, lateral eyes were coded as present only in two groups of Cyphophthalmi (Stylocellidae and Pettalidae) and median eyes were coded as present only in Phalangida (except for troglobitic species, which were scored as absent for all eyes). Second, we implemented a strict coding scheme, where lateral eyes (character 3) were coded as present (state 0) in *P. opilio* and *I. pustulosa*, and as missing data (?) in all other Phalangida; the number of lateral eye pairs (character 4) were coded as one pair (state 4) in *P. opilio* and *I. pustulosa*, and as missing data (?) in all other Phalangida; and the condition of the lateral eye rhabdomes was scored as unknown (?) in all Phalangida. Third, given that the common ancestor of *P. opilio* and *I. pustulosa* is equivalent to the common ancestor of all Phalangida, we trialed a ground plan coding scheme where all non-troglobitic Phalangida were scored as bearing a single pair of lateral eyes, with an unknown condition of the lateral eye rhabdomes. The alignments are available as Supplementary Data 2.

Maximum likelihood analyses were conducted with IQ-TREE (Nguyen et al., 2015), with six partitions (morphology and five loci). Best-fitting one-paramater Markov and GTR+I +G models were inferred for the morphological and molecular data partitions, respectively. Nodal support was inferred using 1000 ultrafast bootstrap resampling replicates.

Total evidence dating was performed in MrBayes v.3.2 (Ronquist and Huelsenbeck, 2003), following previously detailed methods for optimization of priors for an independent gamma rates clock model. Six partitions were implemented (morphology and six loci), with third codon positions of two protein-coding genes excluded to retain only conserved sites. Molecular substitution and clock model parameters were unlinked across partitions. *Limulus polyphemus* (Xiphosura) was set as the outgroup and a uniform prior spanning 500-550 Mya was set for the root age, reflecting recent divergence time estimations of the chelicerate radiation. A one-parameter Markov model (Lewis,2001) was applied to the morphological data. Best-fitting one-parameter Markov and GTR+I +G models were inferred for the morphological and molecular data partitions, respectively. Fossils were implemented as dated tips in the analysis and uncertainty in fossil ages was disregarded, as it was considered negligible with respect to the scale of geological time spanned by Opiliones. Four runs, each with three hot and one cold Markov chains, were executed for 10^7 M generations. Convergence diagnostics were assessed using Tracer v.1.7 (Rambaut et al., 2018) and 25% of each run was discarded as burnin. We implemented two alternative coding strategies for the morphological data: (1) the traditional coding and (2) the strict coding, as described above for the maximum likelihood analyses. Command files are available as Supplementary Data 2.

References

Alberti, G., Lipke, E., Giribet, G. (2008). On the ultrastructure and identity of the eyes of Cyphophthalmi based on a study of *Stylocellus* sp. (Opiliones, Stylocellidae). *J. Arachnol.* 36, 379–387.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Alvarez, C. E. (2008). On the origins of arrestin and rhodopsin. *BMC Evol. Biol.* 8, 222–13.

Asher, R. J., Smith, M. R., Rankin, A. and Emry, R. J. (2019). Congruence, fossils and the evolutionary tree of rodents and lagomorphs. *Roy. Soc. Open Sci.* 6, 190387.

Baird, E. and Yilmaz, A. (2023). Insect Dorsal Ocelli: A Brief Overview. In *Distributed Vision* (ed. Buschbeck, E. and Bok, M.), pp. 205–221. Springer, Cham.

Ballesteros, J. A., Santibáñez-López, C. E., Baker, C. M., Benavides, L. R., Cunha, T. J., Gainett, G., Ontano, A. Z., Setton, E. V. W., Arango, C. P., Gavish-Regev, E., et al. (2022). Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Mol. Biol. Evol.* 39, msac021.

Battelle, B.-A. (2006). The eyes of *Limulus polyphemus* (Xiphosura, Chelicerata) and their afferent and efferent projections. *Arthropod Struct. Dev.* 35, 261–274.

Battelle, B., Dabdoub, A., Malone, M. A., Andrews, A. W., Cacciatore, C., Calman, B. G., Smith, W. C. and Payne, R. (2001). Immunocytochemical localization of opsin, visual arrestin, myosin III, and calmodulin in *Limulus* lateral eye retinular cells and ventral photoreceptors. *J. Comp. Neurol.* 435, 211–225.

Battelle, B.-A., Ryan, J. F., Kempler, K. E., Saraf, S. R., Marten, C. E., Warren, W. C., Minx, P. J., Montague, M. J., Green, P. J., Schmidt, S. A., et al. (2016). Opsin repertoire and expression patterns in horseshoe crabs: Evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biol. Evol.* 8, 1571–1589.

Beck, R. M. D. and Baillie, C. (2018). Improvements in the fossil record may largely resolve current conflicts between morphological and molecular estimates of mammal phylogeny. *Proc. R. Soc. B* 285, 20181632.

Bonini, N. M., Leiserson, W. M. and Benzer, S. (1993). The eyes absent gene: Genetic control of cell survival and differentiation in the developing *Drosophila* eye. *Cell* 72, 379–395.

Brenneis, G. (2022). The visual pathway in sea spiders (Pycnogonida) displays a simple serial layout with similarities to the median eye pathway in horseshoe crabs. *BMC Biol.* 20, 27.

Bruce, H. S., Jerz, G., Kelly, S. R., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D. A., Wolff, C. and Patel, N. H. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol V.1.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *CellReports* 18, 762–776.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.

Cheyette, B. N. R., Green, P. J., Martin, K., Garren, H., Hartenstein, V. and Zipursky, S. L. (1994). The *Drosophila sine oculis* locus encodes a homeodomain-containing protein required for the development of the entire visual system. *Neuron* 12, 977–996.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A. and Pierce, N. A. (2018). Third-generation in situ hybridization chain

reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753.
Cobbett, A., Wilkinson, M., Wills, M. A. and Sullivan, J. (2007). Fossils impact as hard as living taxa in parsimony analyses of morphology. *Systematic Biol.* 56, 753–766.
Cohn, M. J. and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. *Nature* 399, 474–479.
Coiro, M., Chomicki, G. and Doyle, J. A. (2018). Experimental signal dissection and method sensitivity analyses reaffirm the potential of fossils and morphology in the resolution of the relationship of angiosperms and Gnetales. *Paleobiology* 44, 490–510.
Darwin, C. (1859). *On the Origin of Species*. (ed. Murray, J.) London.
Davesne, D., Gallut, C., Barriel, V., Janvier, P., Lecointre, G. and Otero, O. (2016). The phylogenetic intrarelationships of spiny-rayed fishes (Acanthomorpha, Teleostei, Actinopterygii): fossil taxa increase the congruence of morphology with molecular data. *Frontiers Ecol. Evol.* 4, 129.
Doeffinger, C., Hartenstein, V. and Stollewerk, A. (2010). Compartmentalization of the precheliceral neuroectoderm in the spider *Cupiennius salei*: Development of the arcuate body, optic ganglia, and mushroom body. *J. Comp. Neurol.* 518, 2612–2632.
Donoghue, M. J., Doyle, J. A., Gauthier, J., Kluge, A. G. and Rowe, T. (1989). The importance of fossils in phylogeny reconstruction. *Annu. Rev. Ecol. Syst.* 20, 431–460.
Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
Elofsson, R. (2006). The frontal eyes of crustaceans. *Arthropod Struct. Dev.* 35, 275–291.
Fernández, R., Sharma, P. P., Tourinho, A. L. and Giribet, G. (2017). The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. *Proc. R. Soc. B* 284, 20162340.
Friedrich, M. (2006). Ancient mechanisms of visual sense organ development based on comparison of the gene networks controlling larval eye, ocellus, and compound eye specification in *Drosophila*. *Arthropod Struct. Dev.* 35, 357–378.
Gainett, G., Ballesteros, J. A., Kanzler, C. R., Zehms, J. T., Zern, J. M., Aharon, S., Gavish-Regev, E. and Sharma, P. P. (2020). Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genomics* 21, 811–17.
Gainett, G., González, V. L., Ballesteros, J. A., Setton, E. V. W., Baker, C. M., Gargiulo, L. B., Santibáñez-López, C. E., Coddington, J. A. and Sharma, P. P. (2021). The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proc. R. Soc. B* 288, 20211168.
Gainett, G., Crawford, A. R., Klementz, B. C., So, C., Baker, C. M., Setton, E. V. W. and Sharma, P. P. (2022). Eggs to long-legs: embryonic staging of the harvestman *Phalangium opilio* (Opiliones), an emerging model arachnid. *Front. Zool.* 19, 11.
Garwood, R. J., Sharma, P. P., Dunlop, J. A. and Giribet, G. (2014). A Paleozoic stem group to mite harvestmen revealed through integration of phylogenetics and development. *Curr. Biol.* 24, 1017–1023.
Gehring, W. J. (2001). The genetic control of eye development and its implications for the evolution of the various eye-types. *Zoology* 104, 171–183.
Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.

Hall, B. K. (2003). Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biol. Rev.* 78, 409–433.

Harzsch, S., Vilpoux, K., Blackburn, D. C., Platchetzki, D., Brown, N. L., Melzer, R., Kempler, K. E. and Battelle, B. A. (2006). Evolution of arthropod visual systems: development of the eyes and central visual pathways in the horseshoe crab *Limulus polyphemus* Linnaeus, 1758 (Chelicerata, Xiphosura). *Dev. Dyn.* 235, 2641–2655.

Henze, M. J. and Oakley, T. H. (2015). The dynamic evolutionary history of pancrustacean eyes and opsins. *Integr. Comp. Biol.* 55, 830–842.

Koch, N. M. and Parry, L. A. (2020). Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Systematic Biol.* 69, 1052–1067.

Koressaar, T. and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.

Kuehn, E., Clausen, D. S., Null, R. W., Metzger, B. M., Willis, A. D. and Özpolat, B. D. (2022). Segment number threshold determines juvenile onset of germline cluster expansion in *Platynereis dumerilii*. *J. Exp. Zoology Part B Mol. Dev. Evol.* 338, 225–240.

Kumar, J. P. (2009). The molecular circuitry governing retinal determination. *Biochim. Biophys. Acta* 1789, 306–314.

Lan, T., Zhao, Y., Zhao, F., He, Y., Martinez, P. and Strausfeld, N. J. (2021). Leanchoiliidae reveals the ancestral organization of the stem euarthropod brain. *Curr. Biol.* 31, 4397-4404.e2.

Land, M. F. (1985). The Morphology and Optics of Spider Eyes. In *Neurobiology of Arachnids*, pp. 53–78. Springer-Verlag Berlin Heidelberg.

Legg, D. A. (2014). *Sanctacaris uncata*: the oldest chelicerate (Arthropoda). *Naturwissenschaften* 101, 1065–1073.

Lehmann, T. and Melzer, R. R. (2013). Looking like *Limulus*?–Retinula axons and visual neuropils of the median and lateral eyes of scorpions. *Front. Zool.* 10, 40–40.

Lehmann, T., Lodde-Bensch, E., Melzer, R. R. and Metz, M. (2016). The visual system of harvestmen (Opiliones, Arachnida, Chelicerata)–a re-examination. *Front. Zool.* 13, 1–16.

Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D. and Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* 10, 2295–8.

Miether, S. T. and Dunlop, J. A. (2016). Lateral eye evolution in the arachnids. *Arachnology* 17, 103–119.

Morehouse, N. I., Buschbeck, E. K., Zurek, D. B., Steck, M. and Porter, M. L. (2017). Molecular evolution of spider vision: new opportunities, familiar players. *Biol. Bull.* 233, 21–38.

Moritz, M. (1957). Zur Embryonalentwicklung der Phalangiiden (Opiliones, Palpatores) unter besonder Berücksichtigung der äußeren Morphologie, der Bildung des Mitteldarmes und der Genitalanlage. *Zool. Jahrb. Abt. Anat. Ontog. Tiere* 76, 331–370.

Nagata, T., Koyanagi, M., Tsukamoto, H. and Terakita, A. (2009). Identification and characterization of a protostome homologue of peropsin from a jumping spider. *J Comp Physiology* 196, 51.

Nagata, T., Koyanagi, M., Tsukamoto, H., Saeki, S., Isono, K., Shichida, Y., Tokunaga, F., Kinoshita, M., Arikawa, K. and Terakita, A. (2012). Depth perception from image defocus in a jumping spider. *Science* 335, 469–471.

Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.

Ortega-Hernández, J., Lerosey-Aubril, R., Losso, S. R. and Weaver, J. C. (2022). Neuroanatomy in a middle Cambrian mollisoniid and the ancestral nervous system organization of chelicerates. *Nat. Commun.* 13, 410.

Patterson, C. (1981). Significance of fossils in determining evolutionary relationships. *Annu. Rev. Ecol. Syst.* 12, 195–223.

Paulus, H. F. (1979). Eye structure and the monophyly of the Arthropoda. In *Arthropod Phylogeny* (ed. Gupta, A. P.), pp. 299–383. Van Nostrand Reinhold.

Pignoni, F., Hu, B., Zavitz, K. H., Xiao, J., Garrity, P. A. and Zipursky, S. L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development. *Cell* 91, 881–891.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Systematic Biol.* 67, 901–904.

Ramirez, M. D., Pairett, A. N., Pankey, M. S., Serb, J. M., Speiser, D. I., Swafford, A. J. and Oakley, T. H. (2016). The last common ancestor of most bilaterian animals possessed at least 9 opsins. *Genome Biol. Evol.* evw248-13.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.

Samadi, L., Schmid, A. and Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *EvoDevo* 6, 16.

Schoenemann, B. and Clarkson, E. N. K. (2023). The median eyes of trilobites. *Sci. Rep.* 13, 3917.

Schomburg, C., Turetzek, N., Schacht, M. I., Schneider, J., Kirfel, P., Prpic, N.-M. and Posnien, N. (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo* 6, 15.

Sharma, P., Giribet, G. (2006). A new *Pettalus* species (Opiliones, Cyphophthalmi, Pettalidae) from Sri Lanka with a discussion on the evolution of eyes in Cyphophthalmi. *J. Arachnol.* 34, 331–341.

Sharma, P. P. and Giribet, G. (2014). A revised dated phylogeny of the arachnid order Opiliones. *Frontiers Genetics* 5, 255.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012a). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14, 450–463.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012b). Evolution of the chelicera: a *dachshund* domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* 14, 522–533.

Shultz, J. W. and Pinto-da-Rocha, R. (2007). Morphology and functional anatomy. In *Harvestmen: The Biology of Opiliones* (ed. Pinto-da-Rocha, R., Machado, G.), and Giribet, G.), p. Harvard University Press.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.

Smith, W. C., Greenberg, R. M., Calman, B. G., Hendrix, M. M., Hutchinson, L., Donoso, L. A. and Battelle, B. (1995). Isolation and expression of an arrestin cDNA from the horseshoe crab lateral eye. *J. Neurochem.* 64, 1–13.

Strausfeld, N. J., Ma, X., Edgecombe, G. D., Fortey, R. A., Land, M. F., Liu, Y., Cong, P. and Hou, X. (2016). Arthropod eyes: The early Cambrian fossil record and divergent evolution of visual systems. *Arthropod Struct. Dev.*45, 152–172.

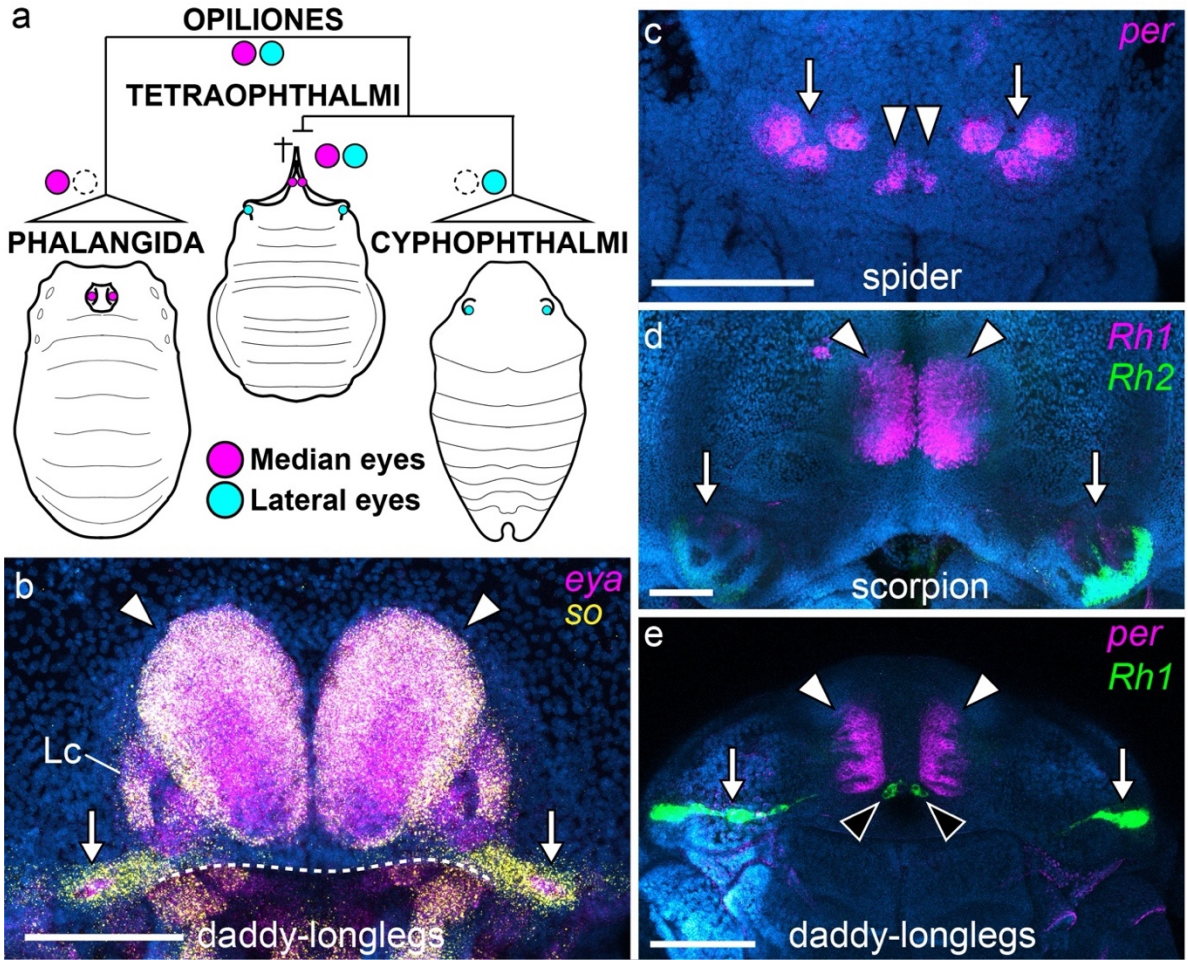


Fig. 1 | The evolution of eyes in harvestmen and presence of rudimentary eyes (molecular markers) in embryos of the daddy-longlegs *P. opilio*. a, Mite harvestmen (Cyphophthalmi) have one pair of lateral eyes, in contrast to their sister group, Phalangida, which have one pair of median eyes. The fossil suborder Tetraophthalmi has both pairs of eyes and was previously recovered as a stem-group Cyphophthalmi (Garwood et al. 2014). **b,** Developing eyes in *P. opilio*, showing expression of *eya* in magenta and *so* in yellow (frontal view). **c,** Head of a scorpion embryo (frontal view). Expression of *Rh1* opsin paralogs is shown in magenta and green. **d,** Head of a spider embryo (frontal view). *peropsin* expression in magenta. **e,** Head of a *P. opilio* daddy-longlegs embryo (frontal view). *peropsin* expression (magenta) and *Rh1* expression (green). Nuclei in blue (Hoechst). White arrow: lateral cells; black arrowhead: median cells; white arrowhead: median eyes. Scale bars: 100 µm.

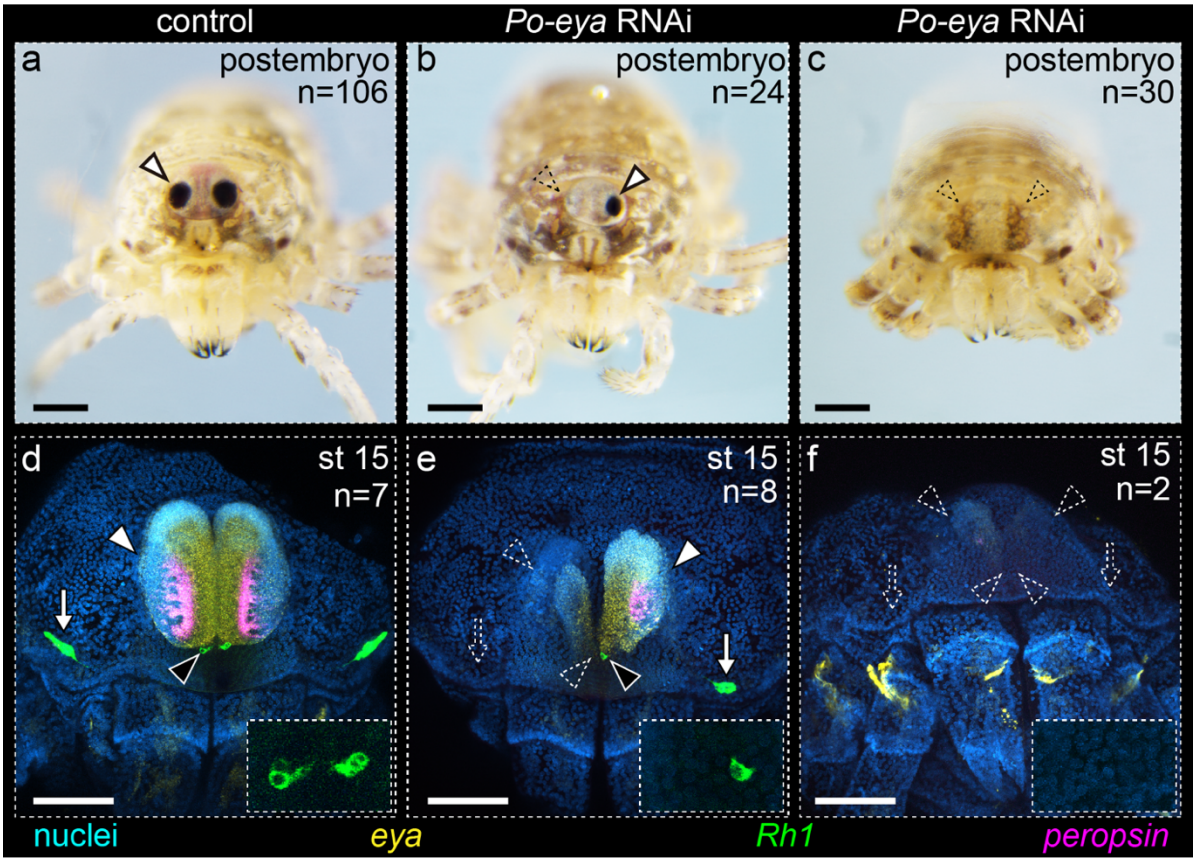


Fig. 2 | Median eyes and rudimentary eyes in *P. opilio* require the Retinal Determination Network gene *eya* for normal development. a–c, frontal view of hatchlings in the control **(a)**, and *eya* RNAi treatment with reduced eyes **(b)** and absent eyes **(c)**. **d–e,** frontal view of stage 15 embryos in the control **(d)** and eyes RNAi treatment **(e–f)**. *eya* (yellow), *Rh1* (green), and *peropsin* (magenta) expression show the loss of both the median eyes and rudimentary eyes upon *eya* knockdown. **(b)** and **(e)** are mosaic embryos. Insets in **(d–f)** show magnification of median cells. Nuclei in blue (Hoechst). White arrow: lateral cells; black arrowhead: median cells; white arrowhead: median eyes. Scale bars: 100 µm.

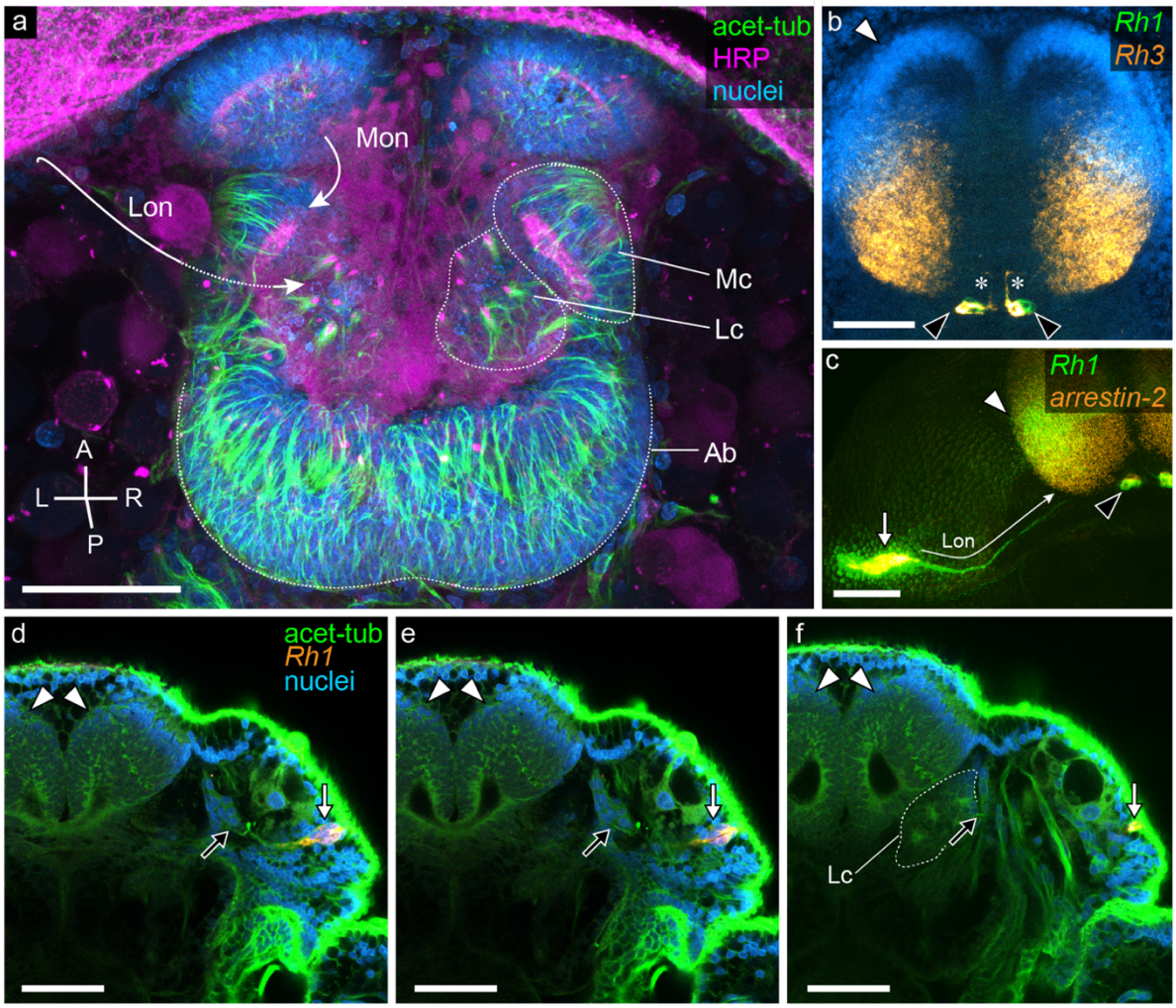


Fig. 3 | *P. opilio* rudimentary eyes are innervated by the protocerebrum. a, Brain of a stage 14/15 *P. opilio* embryo (superior view), indicating the targets of the median eye and rudimentary lateral eye optic nerves (arrows). HRP (magenta) marks subsets of neurons and acetylated tubulin (green) is enriched in neural cells. Validation of eye innervation is shown in Extended Data Figure 8. **b,** *Rh3* (uv-opsin) (orange) and *Rh1* (green) expression mark thin cellular projections from the median cells toward the retinal cells of the median eyes (asterisks). **c,** *arrestin-2* (orange) and *Rh1* expression (green) mark cellular projection of the lateral cells toward the lateral brain center. **d–e,** Serial confocal planes of a frontal view of the embryonic head, from anterior to posterior. *Rh1* (orange) expression marks the lateral cells and acet-tub protein (green) in enriched in neural cells, showing a connection to the lateral brain center (black arrow) (see Supplementary video S3). Nuclei in blue (Hoechst). White arrow: lateral cells; black arrowhead: median cells; white arrowhead: median eyes. Ab: arcuate body; Lon: lateral eye optic nerve; Mon: median eye optic nerve; Mc: Medial brain center; Lc: Lateral brain center. Scale bars: 50 µm.

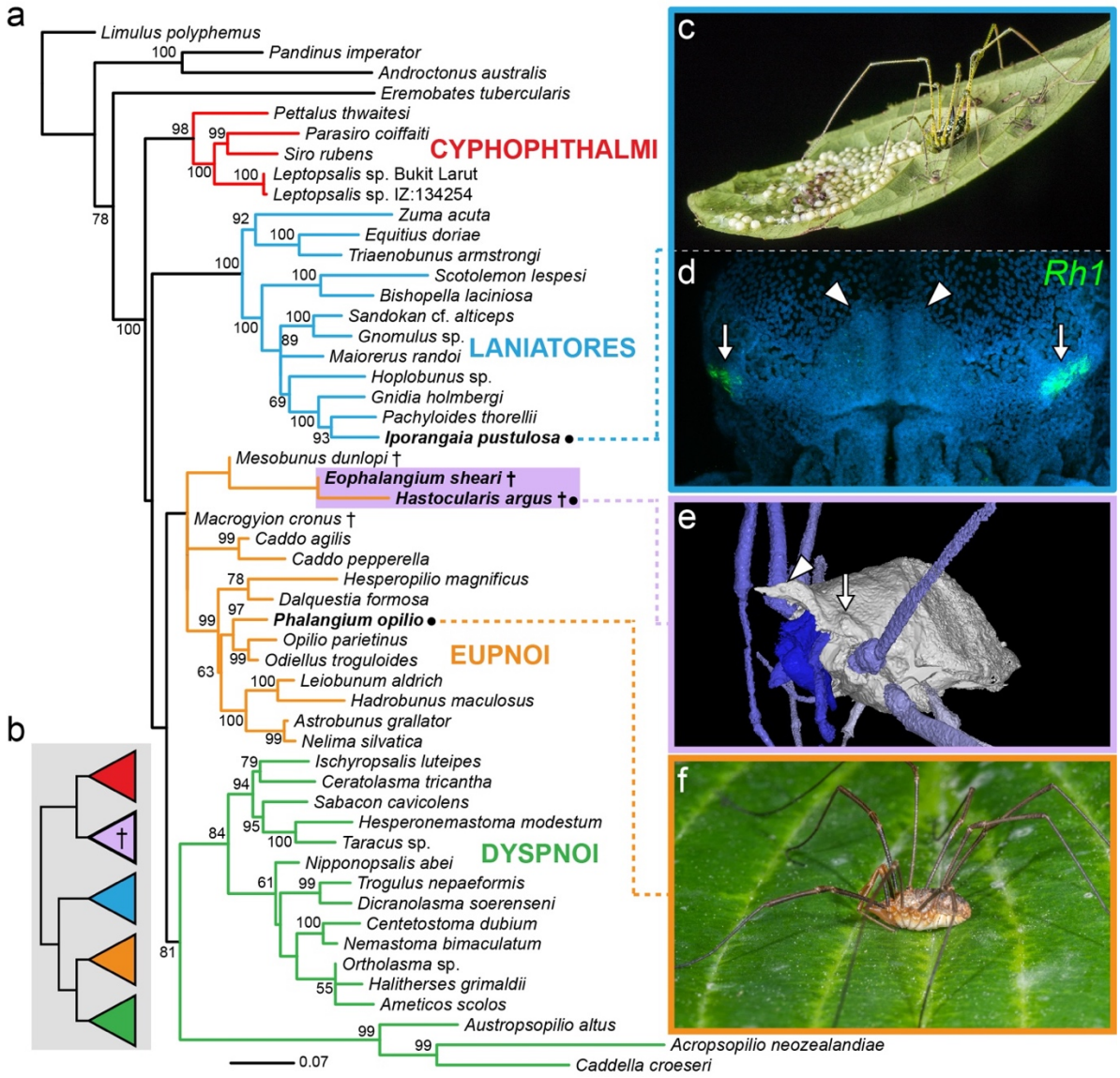
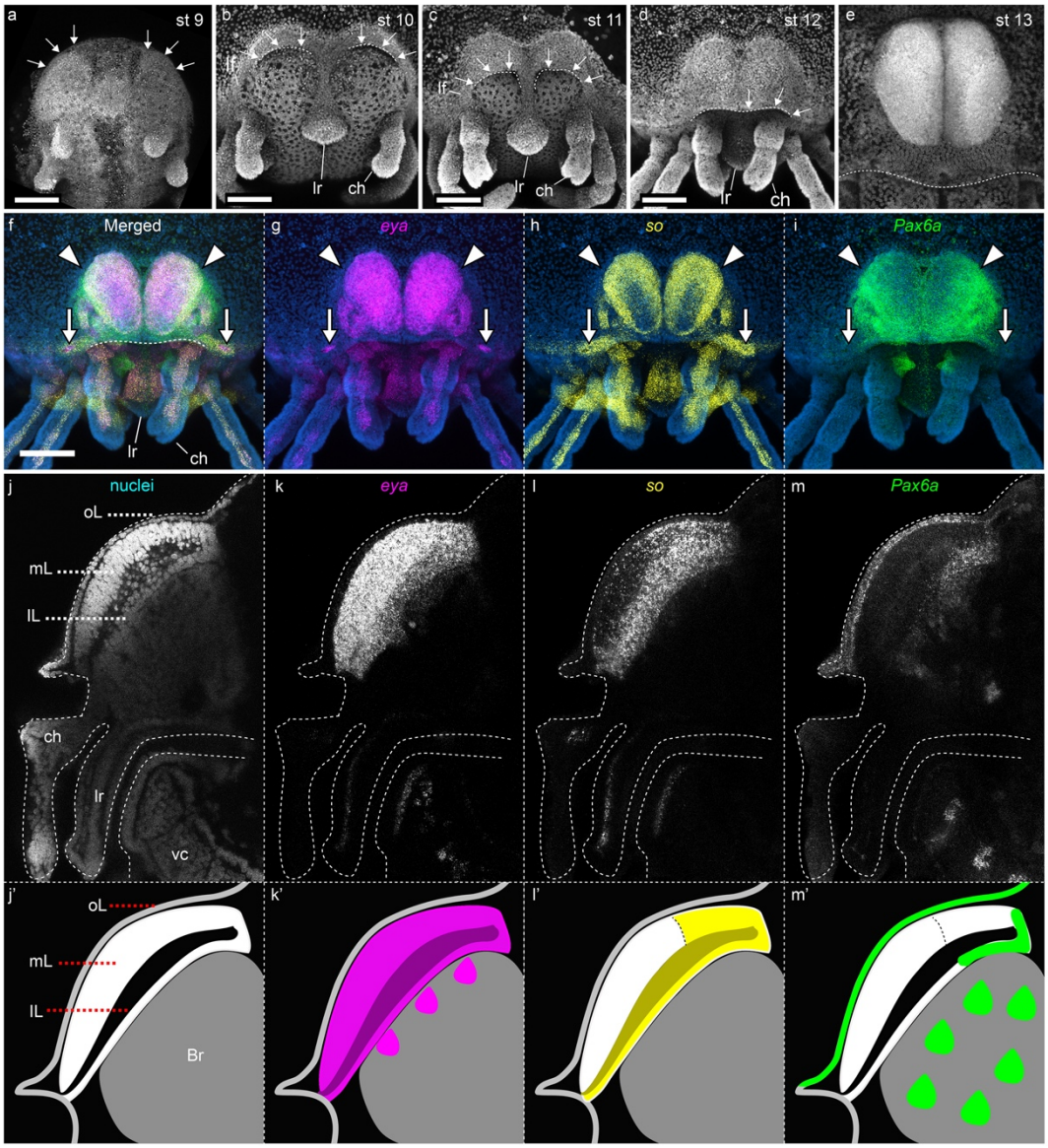
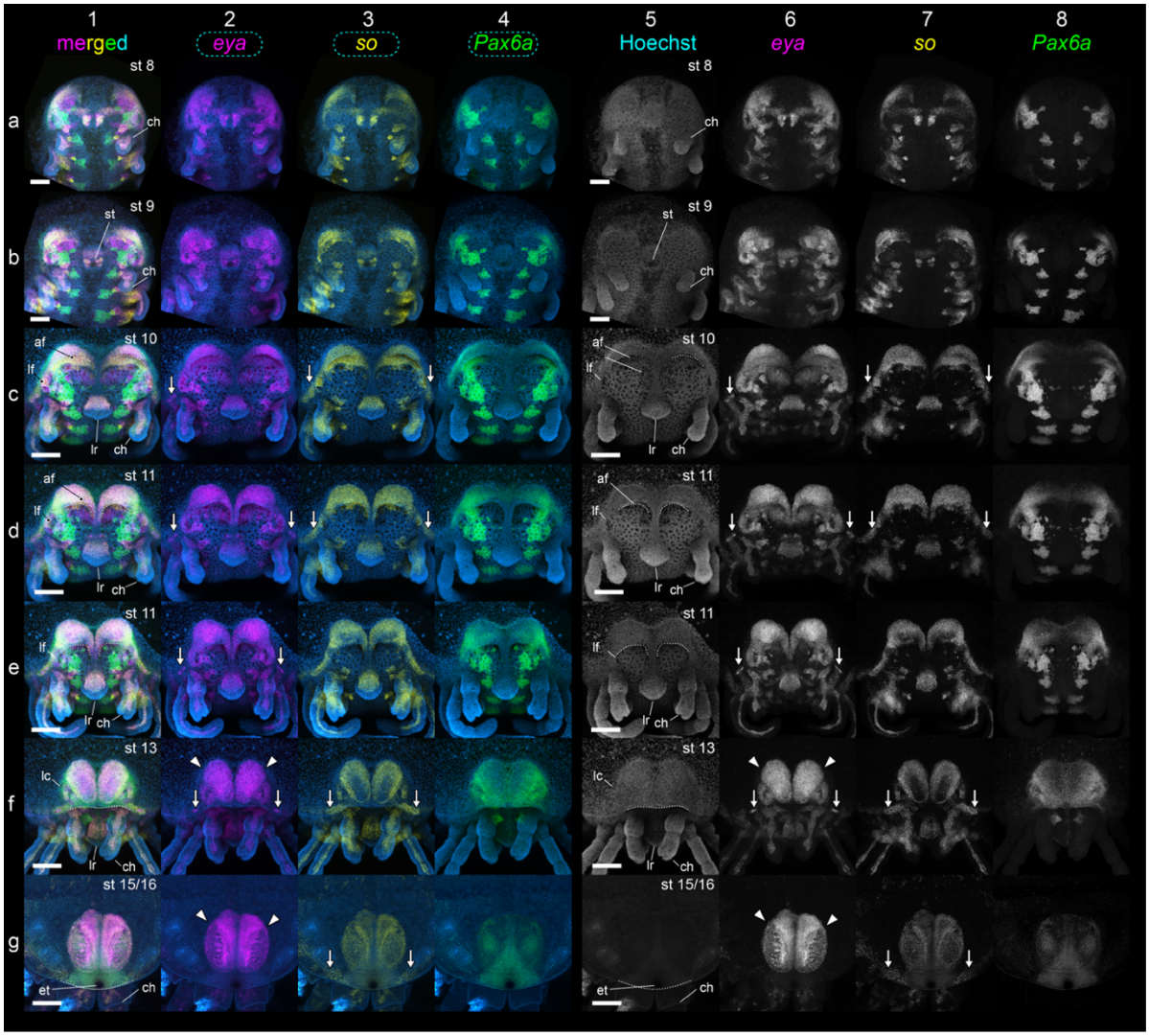


Fig. 4 | *Hastocularis argus*, a four-eyed fossil harvestman, is a crown-group daddy-longlegs. a, Maximum likelihood total evidence topology of Opiliones under the strict coding scheme. Numbers on nodes are ultrafast bootstrap values (values under 55% omitted). **b,** Schematic representation of the position of *H. argus* in the analysis under original coding (lateral eyes missing in Phalangida) (see also Extended Data Fig. 9). **c,** Male armoured harvestman *Iporangaia pustulosa* guarding eggs (photo: John Uribe). **d,** *Rh1* expression in the rudimentary lateral eyes of an embryo of *I. pustulosa*. **e,** Lateral view of *H. argus* 3D reconstruction. **f,** Lateral view of adult male daddy-longlegs *Phalangium opilio* (photo: C.M. Baker).

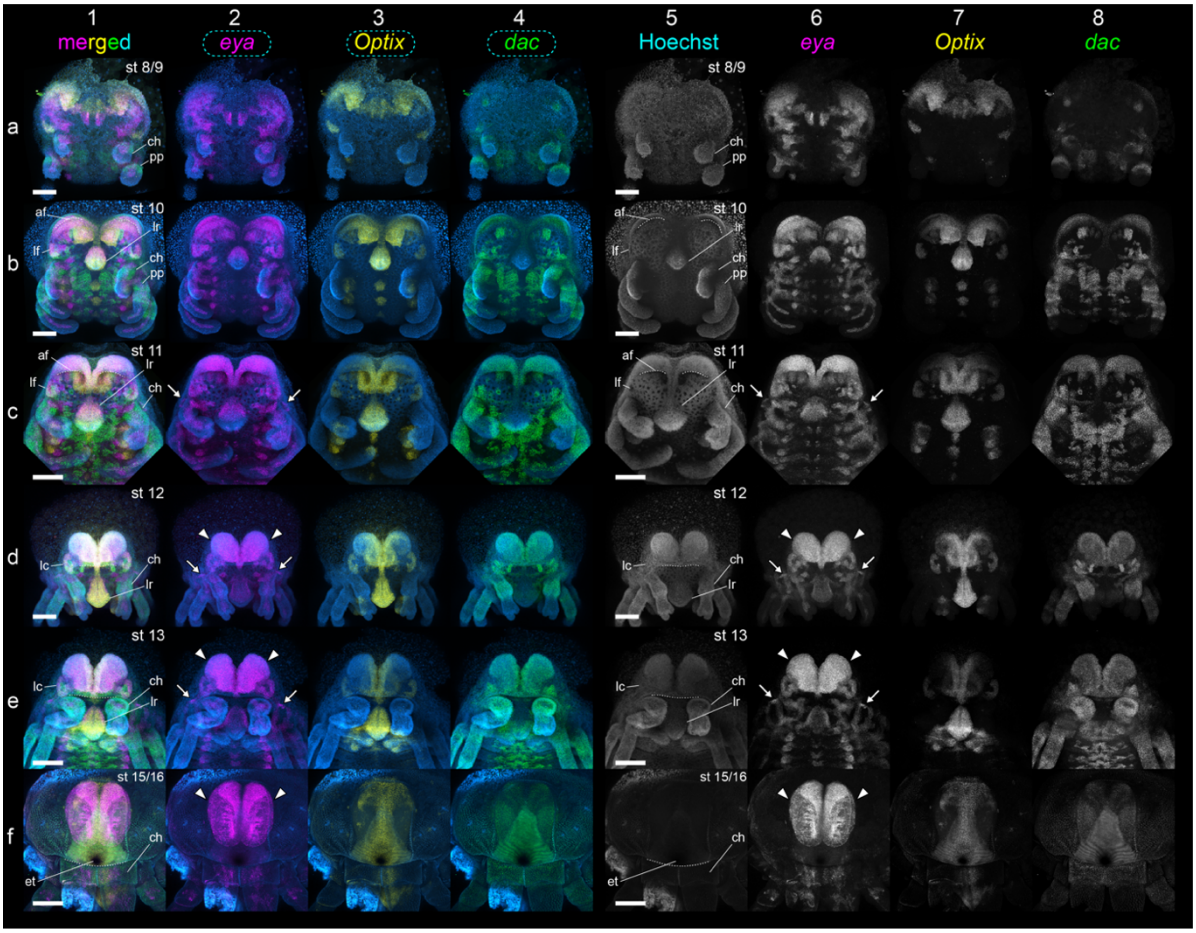
Extended Data Figures and Supplementary information



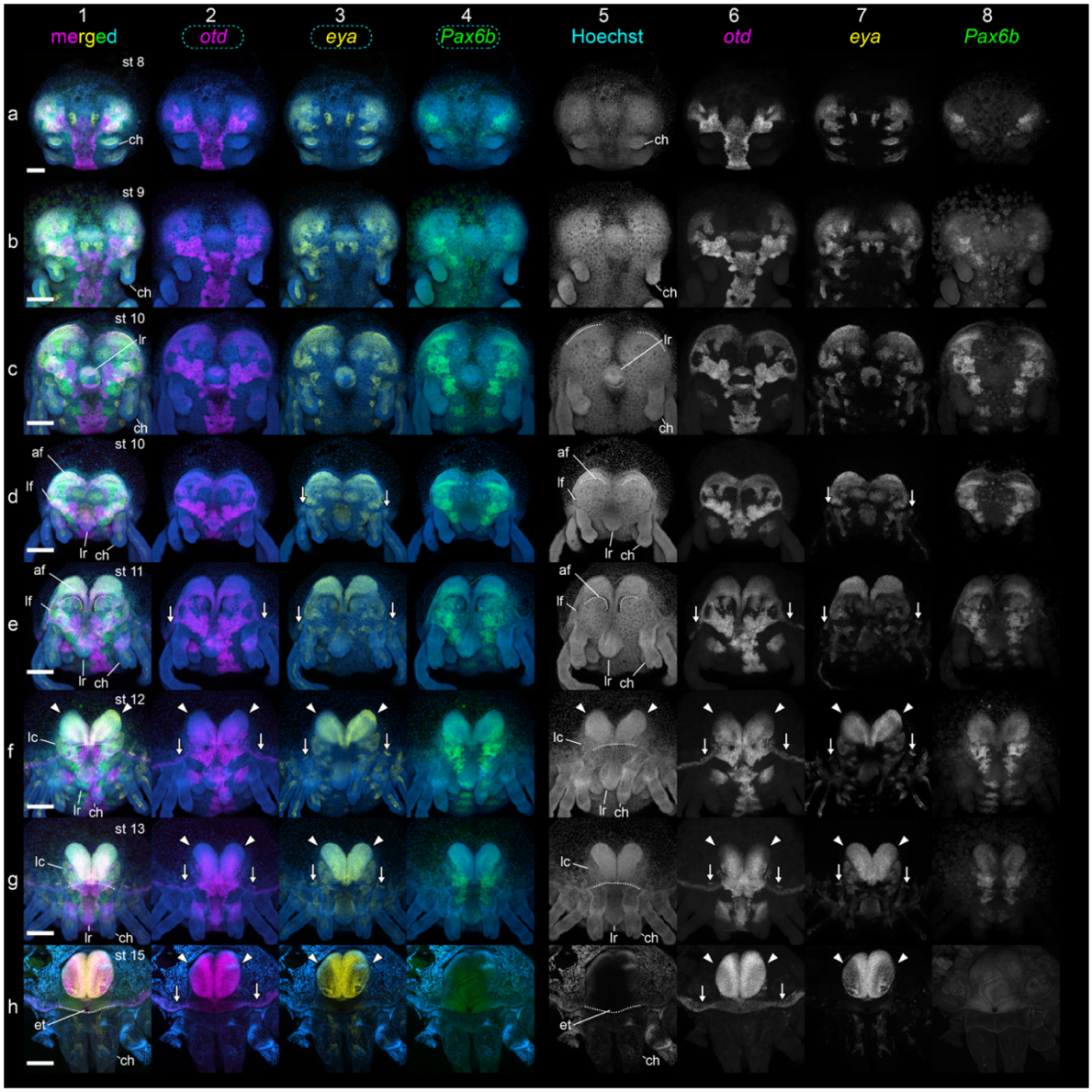
Extended Data Fig. 1 | Eye development in the daddy-longlegs *P. opilio*. a–e, Progression of head development in a series of embryonic stages (Hoechst staining; maximum projection). Arrows mark the anterior margin of the head, which shifts anteriorly to cover the neurogenetic ectoderm. **f–i,** Stage 12 embryo, triple marked with *eya* (magenta), *so* (yellow) and *Pax6a* (green) expression (maximum projection). Nuclei in blue (Hoechst). **j–m,** Single optical slice of the developing median eye (stage 13), triple marked with *eya* (magenta) **(k)**, *so* (yellow) **(i)**, and *Pax6a* (green) **(m)** expression. Note that *Pax6a* is mostly restricted to the outer layer (lentigen layer) of the median eyes, while *eya* and *so* are ubiquitous in the median and inner layers. **j'–m',** Schematic representation of the morphology and expression patterns in the eye cross-section. White arrow: lateral cells; white arrowhead: median eyes. Br: brain; ch: chelicera; ec: ectal compartment of the median eye; lf: lateral furrow; lr: labrum; iL: inner layer of the median eye; mc: mesal compartment of the median eye; mL: median layer of the median eye; oL: outer layer of the median eye; vc: ventral nerve chord. Scale bars: 100 µm.



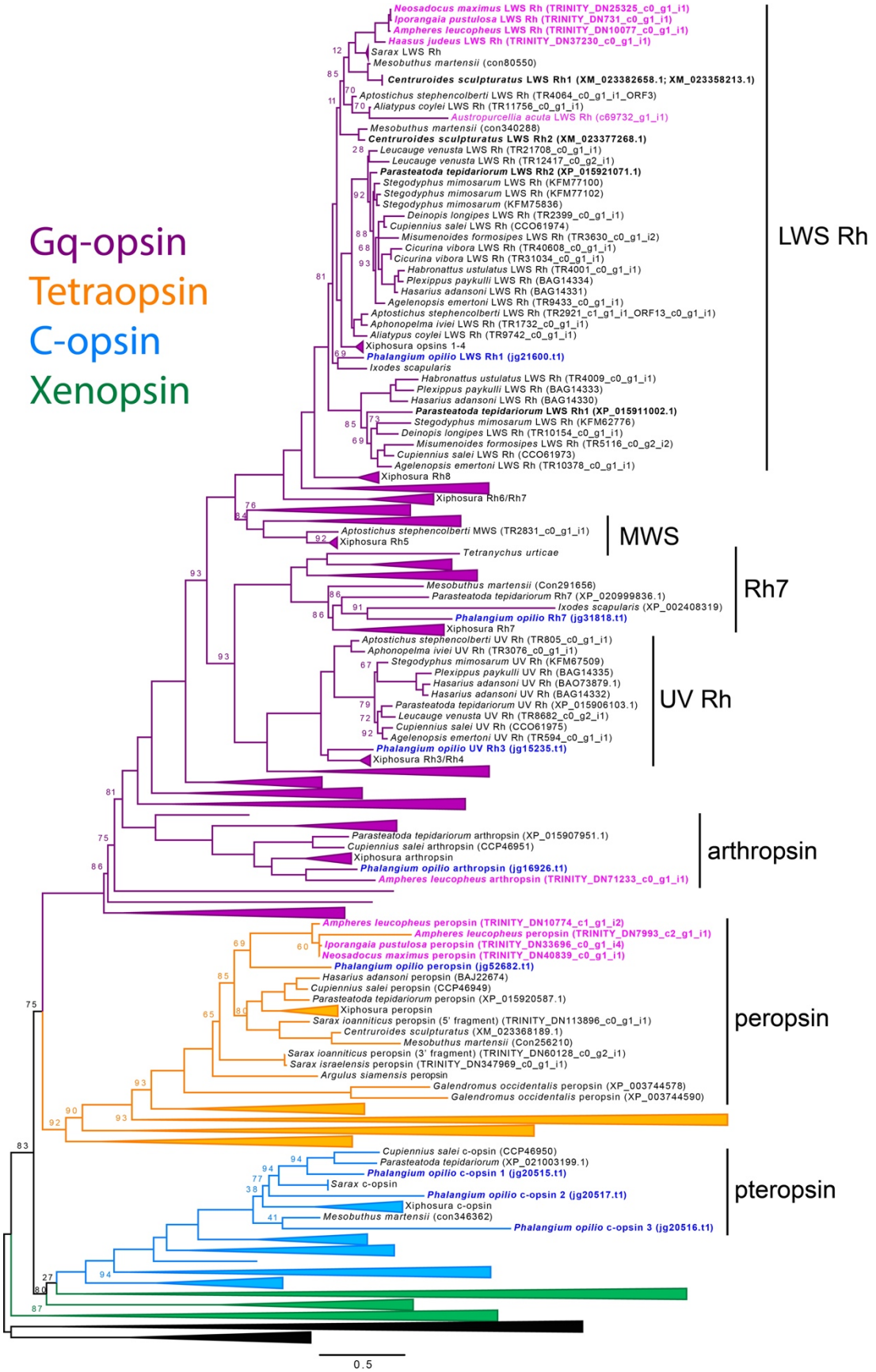
Extended Data Fig. 2 | Atlas of Retinal Determination Network gene expression (maximum projection) across main stages of head development in *P. opilio*. Multiplexed expression patterns of *eyes absent* (*eya*) (magenta), *sine oculis* (*so*) (yellow), and *Pax6a* (green). Frontal view. **a–g,** Each line is a single embryo at different embryonic stages (upper right corner). **Column 1,** Multiplexed expression of the three genes with nuclear counter staining. **Columns 2–4,** single-channel expression (respectively, *eya*, *so* and *Pax6a*), with nuclear counter staining. **Columns 5–8,** Single-channel nuclear staining, and gene expression (respectively, *eya*, *so* and *Pax6a*), in grayscale. White arrow: lateral cells; white arrowhead: median eyes. af: anterior furrow; ch: chelicera; et: egg tooth; lc: lateral brain center; lf: lateral furrow; lr: labrum; st: stomodeum. Scale bars: 100 µm.



Extended Data Fig. 3 | Atlas of Retinal Determination Network gene expression (maximum projection) across main stages of head development in *P. opilio*. Multiplexed expression patterns of *eyes absent* (*eya*) (magenta), *Optix* (yellow), and *dachshund* (*dac*) (green). Frontal view. **a–f,** Each line is a single embryo at different embryonic stages (upper right corner). **Column 1,** Multiplexed expression of the three genes with nuclear counter staining. **Columns 2–4,** single-channel expression (respectively, *eya*, *Optix* and *dac*), with nuclear counter staining. **Columns 5–8,** Single-channel nuclear staining, and gene expression (respectively, *eya*, *Optix* and *dac*), in grayscale. White arrow: lateral cells; white arrowhead: median eyes. af: anterior furrow; ch: chelicera; et: egg tooth; lc: lateral brain center; lf: lateral furrow; lr: labrum. Scale bars: 100 µm.

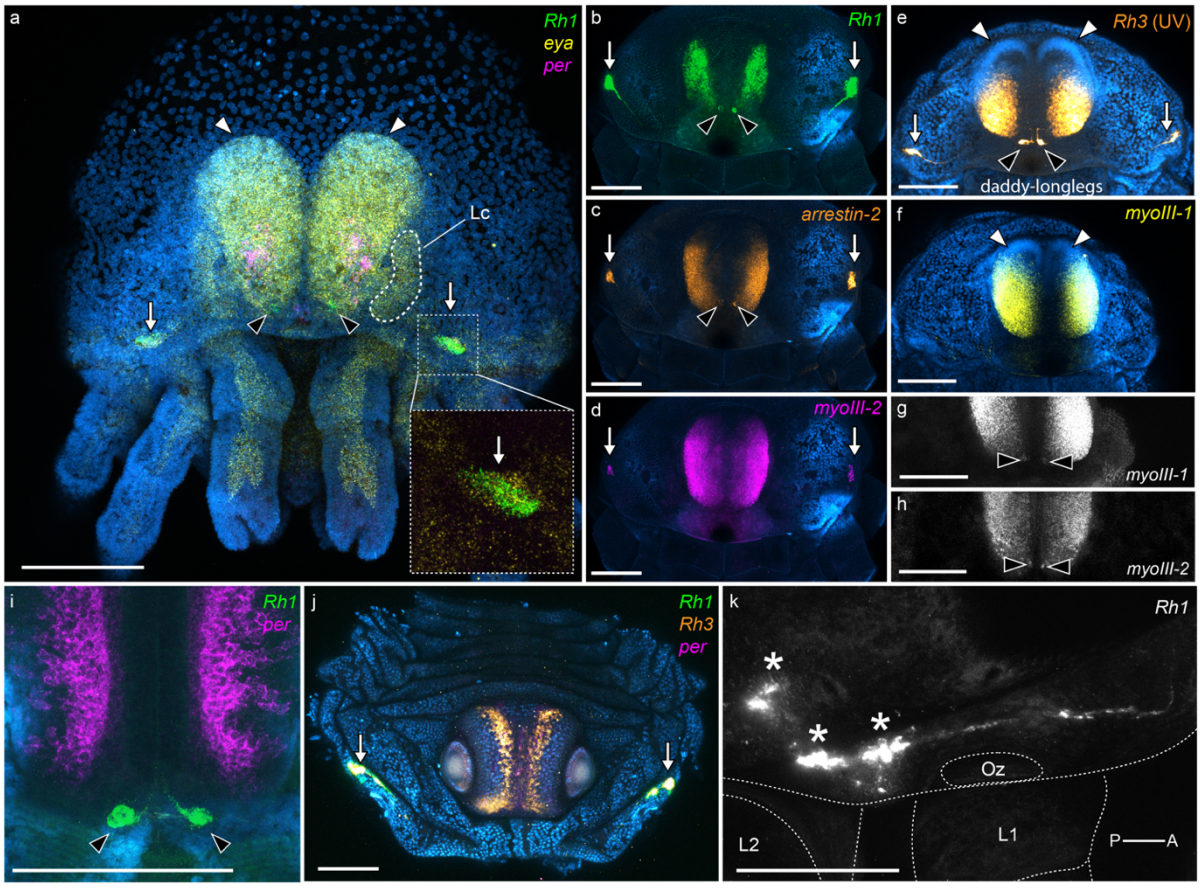


Extended Data Fig. 4 | Atlas of Retinal Determination Network gene expression (maximum projection) across main stages of head development in *P. opilio*. Multiplexed expression patterns of *orthodenticle* (*otd*) (magenta), *eyes absent* (*eya*) (yellow), and *Pax6b* (green). Frontal view. **a–h,** Each line is a single embryo at different embryonic stages (upper right corner). **Column 1,** Multiplexed expression of the three genes with nuclear counter staining. **Columns 2–4,** single-channel expression (respectively, *otd*, *eya* and *Pax6b*), with nuclear counter staining. **Columns 5–8,** Single-channel nuclear staining, and gene expression (respectively, *otd*, *eya* and *Pax6b*), in grayscale. White arrow: lateral cells; white arrowhead: median eyes. af: anterior furrow; ch: chelicera; et: egg tooth; lc: lateral brain center; lf: lateral furrow; lr: labrum. Scale bars: 100 µm.

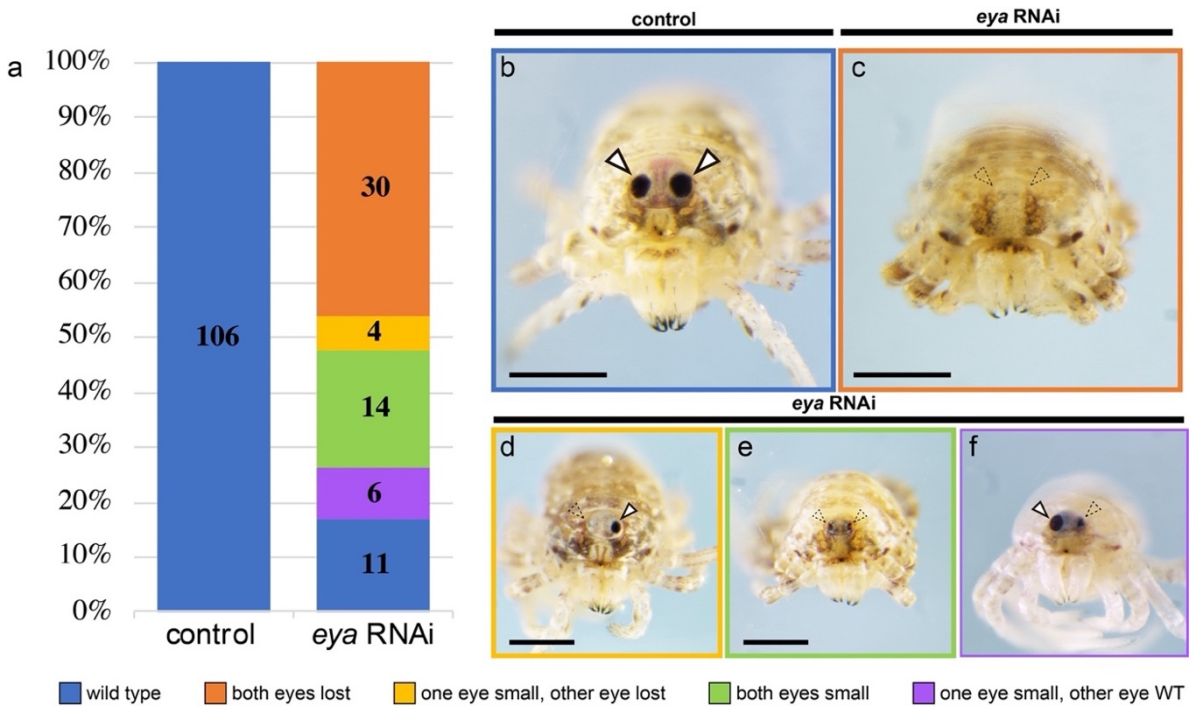


Extended Data Fig. 5 | Tree topology of metazoan opsin genes inferred from maximum likelihood analysis. Dataset compiled after Morehouse et al. (2017). Opiliones terminals

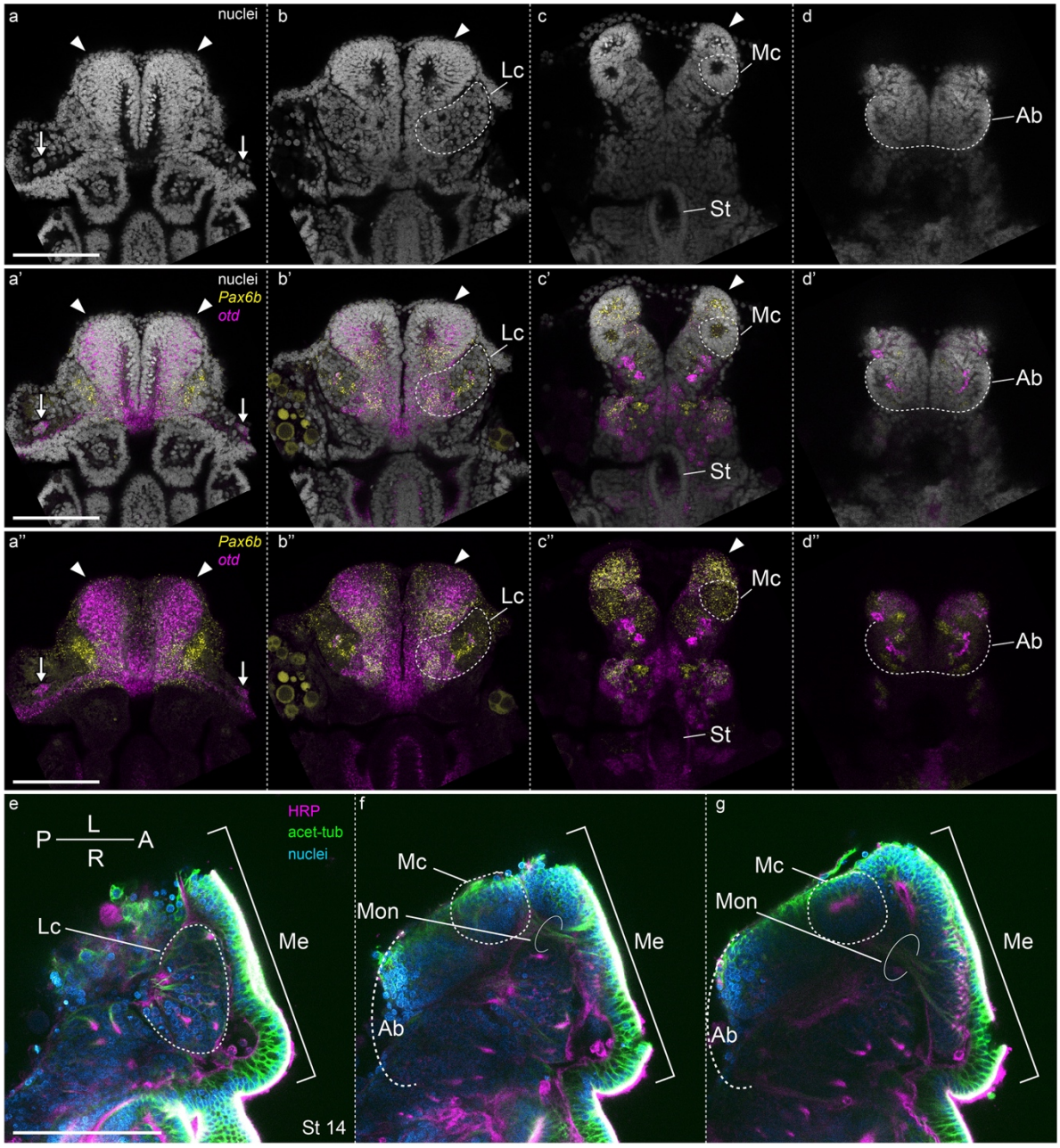
included in this study are in magenta and blue (*P. opilio*). Numbers in nodes are ultrafast bootstrap support values (iqtree -m LG+F+G4 -bb 1000). Major clades and classification of opsins types follow the nomenclature used by Morehouse et al. (2017).



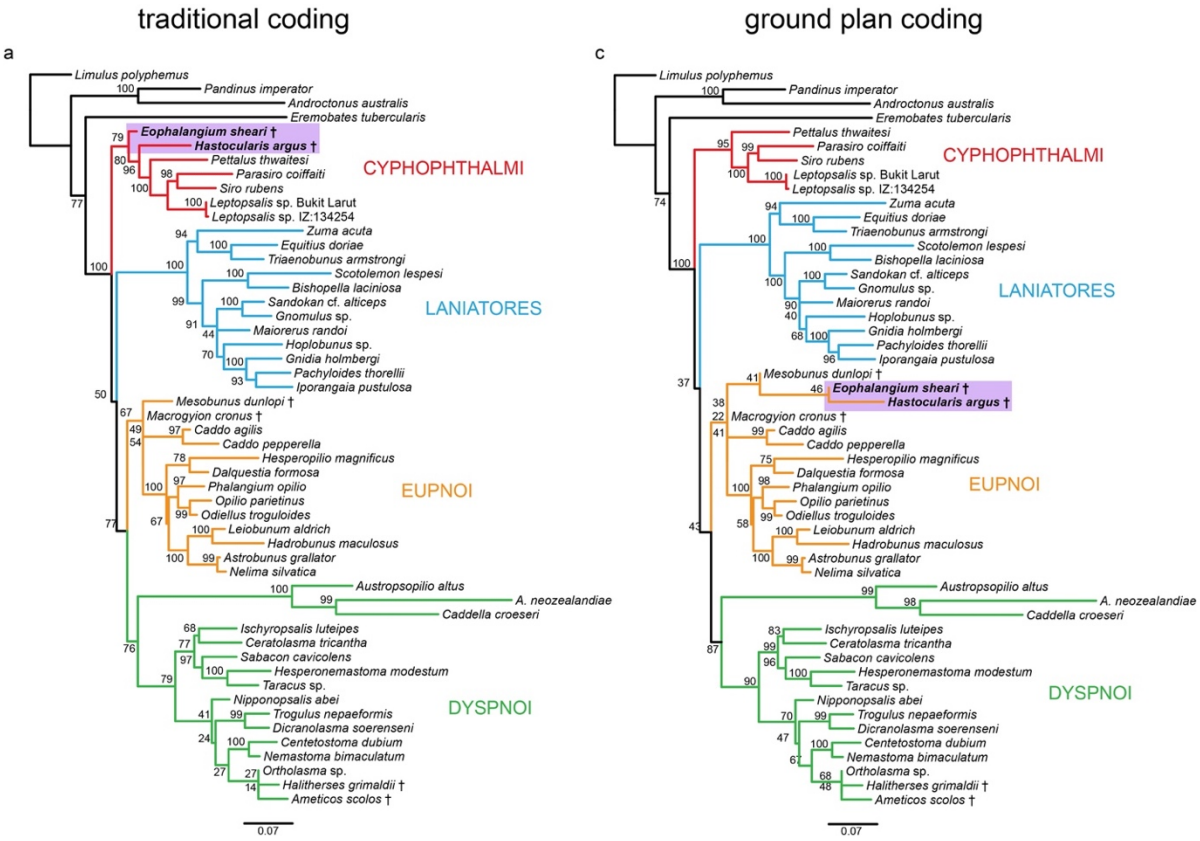
Extended Data Fig. 6 | Median cells and lateral cells express photoreception genes. a, State 13 embryo, showing *eyes absent* (*eya*) (yellow) mRNA co-expression with *Rh1* (green) mRNA in the median and lateral cells, and *peropsin* (magenta) in the median eyes. Inset: detail of the *eya* and *Rh1* expression in the lateral cells. **b–d,** Stage 17 embryo (multiplexed), marked with *Rh1* (green) **(b)**, *arrestin-2* (orange) **(c)**, and *myosin III-2* (*myoIII-2*) (magenta) expression. Note that in this stage *Rh1* is also expressed in the median eyes. **e,** Stage 17, *Rh3* (UV) expression. **f,** Stage17, *myosin III-1* (*myoIII-1*) expression. **g,** sub-stack of **(f)**, showing expression of *myoIII-1* in the median cells. **h,** sub-stack of **(d)**, showing expression of *myoIII-2* in the median cells. **i,** Close-up of *Rh1* expression (green) in the median cells, and *peropsin* expression (magenta) in the median eyes of a stage 16 embryo. **j,** *Rh1* (green), *Rh3* (orange) and *peropsin* (magenta) expression in a third instar individual. **k,** Lateral view of the rudimentary lateral eye in a third instar individual. *Rh1* expression in grey. Asterisk: Concentration of *Rh1* expression; white arrow: lateral cells; black arrowhead: median cells; white arrowhead: median eyes. Lc: lateral brain center; Oz: ozopore; L1–L2: leg 1–2. Scale bars: 100 µm.



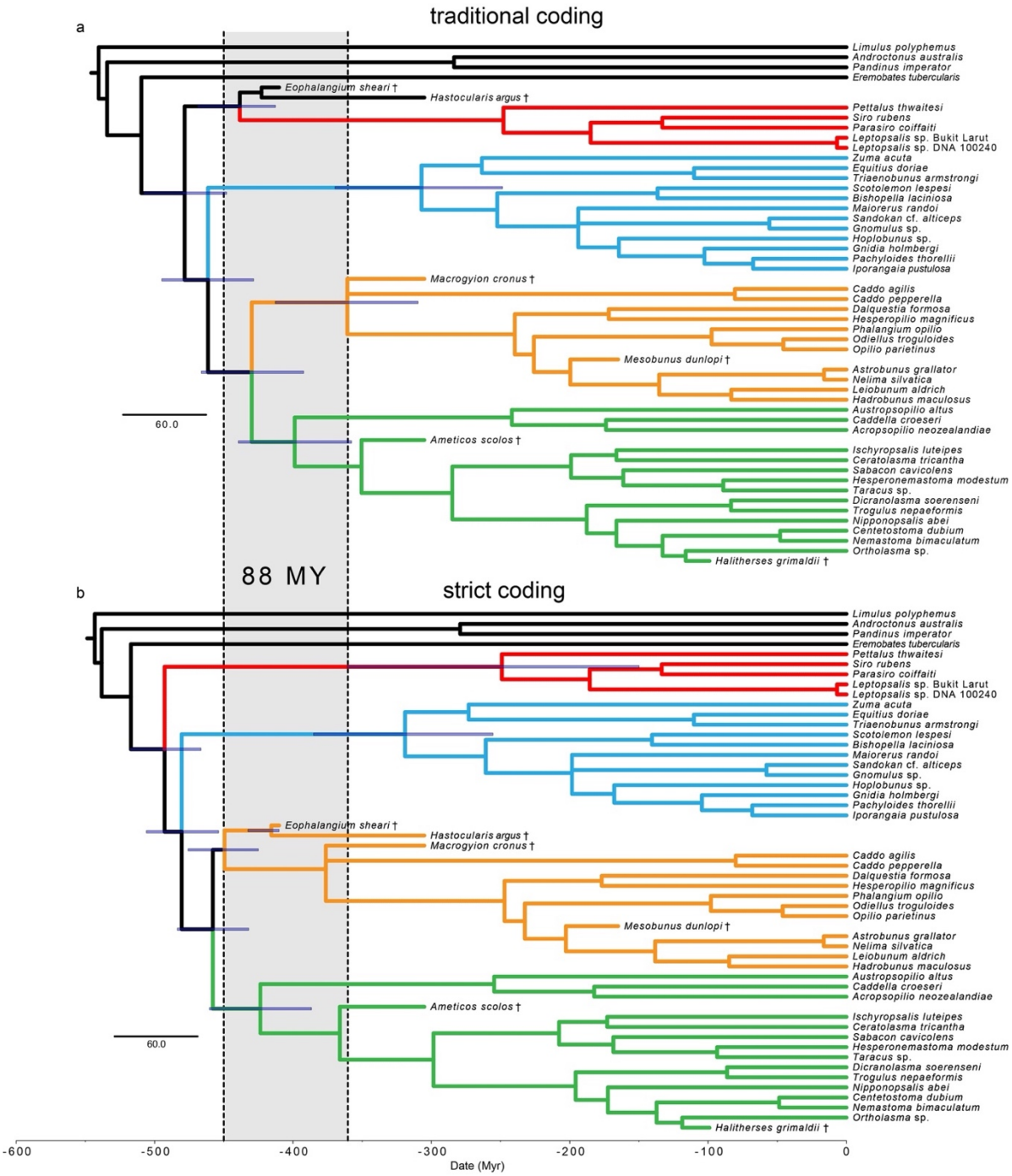
Extended Data Fig. 7 | Summary of *Po-eya* RNA interference (RNAi) gene knockdown experiment. a, Percentage of *P. opilio* hatchlings (y-axis) in the control and *Po-eya* RNAi experiments in the phenotypic categories wild type (blue), both eyes lost (orange), one eye small, other eye lost (yellow), both eyes small (green), and one eye small, other eye wild type (purple). **b,** Frontal view of a wild type hatchling in the control treatment. **c–f:** Frontal view of hatchlings in the *Po-eya* RNAi treatment exemplifying each phenotypic class. **c,** Both eyes lost. **d,** One eye small, other eye lost. **e,** both eyes small. **f,** One eye small, other eye wild type. White arrowhead: median eyes; dashed arrowhead: defective median eye. Scale bars: 100 µm.



Extended Data Fig. 8 | Neuroanatomy of the developing head of *P. opilio*. a–d, Selected Z-stack optical slices of the head in frontal view (stage 14), from anterior to posterior. Nuclei in gray (Hoechst). For the complete Z-stack, see Supplementary Video S1. **a'–d',** Same stack, with nuclei overlayed with *Pax6a* (yellow) and *otd* (magenta) expression. a''–d'': Same stack, showing just *Pax6a* (yellow) and *otd* (magenta) expression. e–f: Selected Z-stack optical slices of the head in lateral view (stage 14). Nuclei in blue, acetylated-tubulin (acet-tub) in green, HRP in magenta. For the complete Z-stack, see Supplementary Video S2. Note the projection from retinal axons towards the median brain center. White arrow: lateral cells; white arrowhead: median eyes. Ab: arcuate body; Lc: lateral brain center; Mc: median brain center; Me: median eye; Mon: median eye optic nerve; St: stomodeum. Scale bars: 100 µm.



Extended Data Fig. 9 | Maximum likehood tree topologies of Opiliones relationships from total evidence analyses under different coding schemes. a, Original coding scheme, scoring lateral eyes as absent in Phalangida (Garwood et al., 2014; Sharma and Giribet, 2014). **b,** Ground plan coding scheme, coding lateral eyes present in all Phalangida terminals (except troglobite species) after the discovery of rudimentary eyes in the present study. Numbers on nodes are ultrafast bootstrap resampling values. Four-eyed fossil suborder Tetraophthalmi is highlighted in purple.



Extended Data Fig. 10 | Total evidence Bayesian dating under alternative coding strategies for the morphological data. a, Traditional coding (lateral eyes absent in Phalangida). **b,** Strict coding (lateral eyes present only in *I. pustulosa* and *P. opilio*). Purple bars depict node age confidence intervals for selected nodes.

Supplementary results

RDN gene expression in the median eyes of *P. opilio*

We refer to the folding head ectoderm from stage 9–11 as the *eye fold* (Extended Data Fig. 1a–c), and to the multilayered eye tissue over the head ectoderm from stage 12 onwards as the *eye field* (Extended Data Fig. 1d–e). The eye field is composed of three layers (Moritz 1957), here referred to as the outer layer, the median layer, and the inner layer (Extended Data Fig. 1j–m'). The eye field has two distinct regions in frontal view, a mesal compartment and an ectal compartment (Extended Data Fig. 1e). All seven genes (*eyes absent* [*Po-eya*], *sine oculis* [*Po-so*], *Po-Optix*, *dachshund* [*Po-dac*], *orthodenticle* [*Po-otd*], and Pax6 paralogs [*Po-Pax6a*; *Po-Pax6b*] are expressed in parts of the developing eye tissue and brain compartments at some stage of development (Extended Data Fig. 2–4). *Po-eya*, *Po-so*, *Po-otd* and *Po-dac* have overlapping expression in the eye tissue throughout eye fold growth and in well-developed eye fields (Extended Data Fig. 1f–h; Extended Data Fig. 2–4), with the exception that *Po-dac* expression becomes restricted to the outer layer of the eye field in late stages (stage 15 onwards) (Extended Data Fig. 3f). *Po-Optix* is expressed in parts of the eye fold, developing brain and labrum, as well as the mesal compartment of the eye field (Extended Data Fig. 3). *Po-Pax6a* and *Po-Pax6b* are mainly expressed in the neural tissue during eye folding, and weakly in the eye field (Extended Data Fig. 1, 4). Eye field layers express different combinations of genes. For instance, while *Po-eya* is ubiquitously expressed in all three layers, *Po-so* is concentrated on the dorsal part of the medium retinal layer, and in the inner retinal layer, whereas *Po-Pax6a* is restricted to the outer lentigen layer (Extended Data Fig.1j–m').

Photoreception gene expression

Po-peropsin is expressed in the inner layer of the ectal compartment the median eyes only, in a ramified pattern (Extended Data Fig. 6i) that overlaps with *Po-eya* transcripts.

We first detected *Po-Rh1* in stage 13/14 in a group of cells at the margin of the prosoma between the coxae of the pedipalp and leg 1 (lateral cells), which coincide with the foci of *Po-eya* expression (Extended Data Fig. 1a). The lateral and central cells continue to develop: the lateral cells extend thin cell processes that could be traced up to the region where the lateral brain center formed. *Po-Rh1* is expressed in the lateral cells and their projections at least until the third instar (out on six to adulthood), when the cuticle starts to become thick and interferes with the in situ hybridization. In the third instar, the lateral cells are concentrated above the coxae of legs 1 and 2, immediately posterior to the ozopore (opening of the repugnatorial gland). Three bright spots of *Po-Rh1* expression can be discerned. We did not detect *Po-arthropsin* expression in the median eyes or lateral cells during the embryonic stages surveyed.

In the laniatorean harvestmen *Iporangaia pustulosa*, LWS Rh1 (*Ip-Rh1*) expression is no longer detected in lateral cells in embryonic stages close to hatching; in these later stages, only the median eyes are associated with *Ip-Rh1* expression.

Supplementary Table 1: List of genes used in the HCR in situ hybridization experiments.

Species	Gene name	Genome ID	Transcriptome ID	
Centruroides sculpturatus	*Rh1*	LOC111632747	NA	
Centruroides sculpturatus	*Rh2*	LOC111616859; LOC111637214	NA	
Centruroides sculpturatus	*peropsin*	LOC111625143	NA	
Iporangaia pustulosa	*Rh1*	NA	TRINITY_DN731_c0_g1_i1	
Iporangaia pustulosa	*peropsin*	NA	TRINITY_DN33696_c0_g1_i4	
Parasteatoda tepidariorum	*Rh1*	LOC107442054	NA	
Parasteatoda tepidariorum	*Rh2*	LOC107449911	NA	
Parasteatoda tepidariorum	*peropsin*	LOC107449547	NA	
Phalangium opilio	*sine oculis/Six1*	jg43040.t1	Popi1_comp39452_c0_seq4	Popi2_comp183481_c0_seq5
		jg130.t1		
Phalangium opilio	*eyes absent*	*jg4550.t1*	*Popi1_comp36689_c0_seq4*	*Popi2_comp166318_c0_seq8*
		jg4550.t2		
Phalangium opilio	*Optix/Six3*	jg30497.t1	Popi1_comp42503_c0_seq3	Popi2_comp165591_c0_seq2
			Popi1_comp42503_c1_seq14	
Phalangium opilio	*Pax6a*	jg4543.t1	*Popi1_comp186280*	
Phalangium opilio	*Pax6b*	jg4545.t1	*Popi1_comp42716_c1_seq5*	
Phalangium opilio	*dachshund*	not predicted	Popi1_comp32064_c0_seq1	Popi2_comp172713_c5_seq1
			Popi1_comp37387_c1_seq1	Popi2_comp28242_c0_seq1
				Popi2_comp186537_c2_seq1
Phalangium opilio	*orthodenticle*	jg181.t1	Popi1_comp42066_c0_seq1	Popi2_comp185461_c0_seq13
Phalangium opilio	*peropsin*	jg52682.t1	not searched	
Phalangium opilio	*Rh1*	jg21600.t1	not searched	
Phalangium opilio	*arthropsin*	jg16926.t1	not searched	
Phalangium opilio	*Rh3*	jg15235.t1	not searched	
Phalangium opilio	*Rh7*	jg31818.t1	not searched	
Phalangium opilio	*myoIII-1*	jg28937.t1	not searched	
Phalangium opilio	*myoIII-2*	jg49605.t1	not searched	
Phalangium opilio	*arrestin-2*	jg47007.t1	not searched	
Phalangium opilio	*arrestin-2-like*	jg51432.t1	not searched	

Supplementary Table 2: List of species and accession numbers of arachnid terminals included in the opsin phylogenetic analysis with metazoan opsins compiled by Morehouse et al (2017).

Species	Gene	Gene ID	Database type	Database ID
Ampheres leucopheus	*Rh1*	TRINITY_DN10077_c0_g1_i1	transcriptome	NCBI (tbd)
Ampheres leucopheus	*peropsin*	TRINITY_DN10774_c1_g1_i2	transcriptome	NCBI (tbd)
Ampheres leucopheus	*peropsin* (fragment)	TRINITY_DN7993_c2_g1_i1	transcriptome	NCBI (tbd)
Ampheres leucopheus	*arthropsin*	TRINITY_DN71233_c0_g1_i1	transcriptome	NCBI (tbd)
Austropurcellia acuta	*Rh1*	c69732_g1_i1	transcriptome	SRX6374210
Centruroides sculpturatus	*Rh1*	LOC111637214	genome	GCA_000671375.2
Centruroides sculpturatus	*Rh1*	LOC111616859	genome	GCA_000671375.2
Centruroides sculpturatus	*Rh2*	LOC111632747	genome	GCA_000671375.2
Centruroides sculpturatus	*peropsin*	LOC111625143	genome	GCA_000671375.2
Iporangaia pustulosa	*Rh1*	TRINITY_DN731_c0_g1_i1	transcriptome	NCBI (tbd)
Iporangaia pustulosa	*peropsin*	TRINITY_DN33696_c0_g1_i4	transcriptome	NCBI (tbd)
Neosadocus maximus	*Rh1*	TRINITY_DN25325_c0_g1_i1	transcriptome	NCBI (tbd)
Neosadocus maximus	*peropsin*	TRINITY_DN40839_c0_g1_i1	transcriptome	NCBI (tbd)
Phalangium opilio	*Rh1*	jg21600.t1	genome	GCA_019434445.1
Phalangium opilio	*Rh3*	jg15235.t1	genome	GCA_019434445.1
Phalangium opilio	*Rh7*	jg31818.t1	genome	GCA_019434445.1
Phalangium opilio	*arthropsin*	jg16926.t1	genome	GCA_019434445.1
Phalangium opilio	*peropsin*	jg52682.t1	genome	GCA_019434445.1
Phalangium opilio	*c-opsin 1*	jg20515.t1	genome	GCA_019434445.1
Phalangium opilio	*c-opsin 2*	jg20517.t1	genome	GCA_019434445.1
Phalangium opilio	*c-opsin 3*	jg20516.t1	genome	GCA_019434445.1

Supplementary Table 3: Primer and cloning detail for *Po-eyes absent*.

Primer ID	Forward primer	Reverse primer	Length	Sequencing ID	Direction	Plasmid sequence	Plasmid #	Comments
Popi_eya	ggccgcggTCTCCACCCG GTGTTAAGTC	cccggggcATGT AACGCCGCCA AATCAG	921	AXW463	Forward	ccaccgacttccaactctagaaaaagtaattca cagtcgggcagaggtcgaggtagaagacaa caaacgaatacgacaacccccgatccggaaa ataacttagaaagggttttcgtttgggatttgga tgaaaccatcgtcgtctttcatactttactgacc ggaacttttgcaacgagattcggaaaggacgt ggctttggggagcgctatgggacataccatg gaacaactcatctaccatttggccgacaatca actatttttcaacgacattgaagattatgatcaa gtacatattgatgacgttttgtccgacgataatg ggcaagatctcagcaattacaatttcgcggcg gatggtttccaagctgcggcttcgaacgcgtc actttgcttaccgacgggagttcggggaggtg tggattggatgaggaaattg	Popi_eya #2	Sequenced fragment is partial; plasmid band matches expected size
Popi_eya	ggccgcggTCTCCACCCG GTGTTAAGTC	cccggggcATGT AACGCCGCCA AATCAG	921	AXW468	Forward	ccaccgacttccaactctagaaaaagtaattc nnantcgggcagaggtcgaggtagaagaca acaaacgaatacgacaacccccgatccggaa aataacttagaaagggttttcgtttgggatttgg atgaaaccatcgtcgtctttcatactttactgac cggaacttttgcaacgagattcggaaaggac gtggctttggggagcgctatgggacataccat ggaacaactcatctaccatttggccgacaatc aactatttttcaacgacattgaagattgtgatca agtacatattgatgacgttttgtccgacgataat gggcaagatctcagcaattacaatttcgcggc gggtggtttccaagctgcggcttcgaacgcgt cactttgcttaccgacgggagttcggggaggt gtggattggatgaggaaattg	Popi_eya #5	Sequenced fragment is partial; plasmid band matches expected size

Supplementary Table 4: List of loci and accession numbers of newly incorporated terminals in the molecular matrix for the total evidence analyses.

Locus	Species	Accession
16S rRNA	*Phalangium opilio*	NC_010766.1
16S rRNA	*Leiobunum aldrichi*	JQ432344.1
16S rRNA	*Hadrobunus maculosus*	JQ432360.1
16S rRNA	*Caddo agilis*	KF963299.1
16S rRNA	*Caddo pepperella*	KF963300.1
16S rRNA	*Eremobates tuberculatus*	KT276604.1
16S rRNA	*Pachyloides thorelli*	KF726570.1
16S rRNA	*Iporangaia pustulosa*	KF726565.1
16S rRNA	*Opilio parietinus*	NC_014700.1
16S rRNA	*Limulus polyphemus*	JX983598.1
28S rRNA	*Zuma acuta*	HM056651.1
28S rRNA	*Dicranolasma soerenseni*	GQ466288.1
28S rRNA	*Astrobunus grallator*	JQ437119.1
28S rRNA	*Nelima silvatica*	JQ437120.1
28S rRNA	*Leiobunum aldrichi*	JQ432286.1
28S rRNA	*Hadrobunus maculosus*	JQ432304.1
28S rRNA	*Dalquestia formosa*	AF124964.1
28S rRNA	*Phalangium opilio*	KJ871594.1
28S rRNA	*Caddo agilis*	KF955597.1
28S rRNA	*Pachyloides thorelli*	KF726682.1
28S rRNA	*Iporangaia pustulosa*	pending
28S rRNA	*Androctonus australis*	pending
18S rRNA	*Iporangaia pustulosa*	pending
cytochrome c oxidase subunit I	*Phalangium opilio*	NC_010766.1
cytochrome c oxidase subunit I	*Opilio parietinus*	NC_014700.1
cytochrome c oxidase subunit I	*Nelima silvatica*	KY270309.1
cytochrome c oxidase subunit I	*Leiobunum aldrichi*	MG510848.1
cytochrome c oxidase subunit I	*Caddo aglis*	MF816271.1
cytochrome c oxidase subunit I	*Ischyropsalis luteipes*	KP224391.1
cytochrome c oxidase subunit I	*Sabacon cavicolens*	KY091837.1
cytochrome c oxidase subunit I	*Taracus* sp.	JX573680.1
cytochrome c oxidase subunit I	*Ortholasma* sp.	GQ912870.1
cytochrome c oxidase subunit I	*Nemastoma bimaculatum*	KY268526.1
cytochrome c oxidase subunit I	*Zuma acuta*	EU162817.1
cytochrome c oxidase subunit I	*Eremobates tuberculatus*	KT276687.1
cytochrome c oxidase subunit I	*Androctonus australis*	MN078204.1
cytochrome c oxidase subunit I	*Iporangaia pustulosa*	KF726789.1
histone H3	*Iporangaia pustulosa*	MG769441.1

Appendix 1

The anatomy of an unstable node: A Levantine relict precipitates phylogenomic dissolution of higher-level relationships of the armored harvestmen (Arachnida: Opiliones: Laniatores)

Aharon, S., Ballesteros, J. A., Crawford, A. R., ***Gainett, G.****, Friske, K., Langford, B., Santibáñez López, C. E., Ya'aran, S., Gavish-Regev, E., Sharma, P. P.*

Published on 3 October 2019 in Invertebrate Systematics

Invertebrate Systematics 33 (5), 697–717.
See attached publication.

Contribution: I conducted field work in Israel that collected the holotype and paratypes series of the new species *Haasus naasane* used for its formal taxonomic description and molecular phylogenetic analysis and revised the manuscript.

Abstract:

After tumultuous revisions to the family-level systematics of Laniatores (the armored harvestmen), the basally branching family Phalangodidae presently bears a disjunct and irregular distribution, attributed to the fragmentation of Pangea. One of the curious lineages assigned to Phalangodidae is the monotypic Israeli genus *Haasus*, the only Laniatores species that occurs in Israel, and whose presence in the Levant has been inferred to result from biogeographic connectivity with Eurasia. Recent surveys of Israeli caves have also yielded a new troglobitic morphospecies of *Haasus*. Here, we describe this new species as *Haasus naasane* sp. nov. So as to test the biogeographic affinity of *Haasus*, we sequenced DNA from both species and RNA from *Haasus naasane* sp. nov., to assess their phylogenetic placement. Our results showed that the new species is clearly closely related to *Haasus judaeus*, but *Haasus* itself is unambiguously nested within the largely Afrotropical family Pyramidopidae. In addition, the Japanese 'phalangodid' *Proscotolemon sauteri* was recovered as nested within the Southeast Asian family Petrobunidae. Phylogenomic placement of *Haasus naasane* sp. nov. in a 1550-locus matrix indicates that Pyramidopidae has an unstable position in the tree of Laniatores, with alternative partitioning of the matrix recovering high nodal support for mutually exclusive tree topologies. Exploration of phylogenetic signal showed the cause of this instability to be a considerable conflict between partitions, suggesting that the basal phylogeny of Laniatores may not yet be stable to addition of taxa. We transfer *Haasus* to Pyramidopidae (new familial assignment). Additionally, we transfer *Proscotolemon* to the family Petrobunidae (new familial assignment). Future studies on basal Laniatores phylogeny should emphasise the investigation of small-bodied and obscure groups that superficially resemble Phalangodidae.

Appendix 2

Phylogenomic resolution of sea spider diversification through integration of multiple data classes

Ballesteros, J. A., *Setton, E. V. W, *Santibáñez López, C. E, Arango, C. P., Brenneis, G., Brix, S., Cano-Sánchez, E., Dandouch, M., Dilly, G. F., **Gainett, G., McAtee, S., McIntyre, L., Moran, A. R., Moran, R., López-González, P., Williamson, C., Woods, H. A., Wheeler, W. C., Sharma, P. P.*

(*co-first author)

Published on 11 September 2021 in Molecular Biology and Evolution

Molecular Biology and Evolution, 38 (2), 686–701.

See attached publication.

Contribution: I performed fieldwork and contributed with samples for DNA extractions and revised the manuscript.

Abstract:

Despite significant advances in invertebrate phylogenomics over the past decade, the higher-level phylogeny of Pycnogonida (sea spiders) remains elusive. Due to the inaccessibility of some small-bodied lineages, few phylogenetic studies have sampled all sea spider families. Previous efforts based on a handful of genes have yielded unstable tree topologies. Here, we inferred the relationships of 89 sea spider species using targeted capture of the mitochondrial genome, 56 conserved exons, 101 ultraconserved elements, and 3 nuclear ribosomal genes. We inferred molecular divergence times by integrating morphological data for fossil species to calibrate 15 nodes in the arthropod tree of life. This integration of data classes resolved the basal topology of sea spiders with high support. The enigmatic family Austrodecidae was resolved as the sister group to the remaining Pycnogonida and the small-bodied family Rhynchothoracidae as the sister group of the robust-bodied family Pycnogonidae. Molecular divergence time estimation recovered a basal divergence of crown group sea spiders in the Ordovician. Comparison of diversification dynamics with other marine invertebrate taxa that originated in the Paleozoic suggests that sea spiders and some crustacean groups exhibit resilience to mass extinction episodes, relative to mollusk and echinoderm lineages.

Appendix 3

Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions

Ontano, A. Z., ***Gainett, G.****, Aharon, S., Ballesteros, J. A., Benavides, L. R., Corbett, K. F., Gavish-Regev, E., Harvey, M. S., Monsma, S., Santibáñez-López, C. E., Setton, E. V. W., Zehms, J. T., Zeh, J. A., Zeh, D. W., Sharma, P. P.*

Published on 10 February 2021 in Molecular Biology and Evolution

Molecular Biology and Evolution, 36 (6), 2446–2467.

See attached publication.

Contribution: I contributed the transcriptomic assemblies of three Amblypygi species, conducted part of the bioinformatic analyses assessing the presence of duplications in Homeobox genes across arthropods and revised the manuscript.

Abstract:

Long-branch attraction is a systematic artifact that results in erroneous groupings of fast-evolving taxa. The combination of short, deep internodes in tandem with long-branch attraction artifacts has produced empirically intractable parts of the Tree of Life. One such group is the arthropod subphylum Chelicerata, whose backbone phylogeny has remained unstable despite improvements in phylogenetic methods and genome-scale data sets. Pseudoscorpion placement is particularly variable across data sets and analytical frameworks, with this group either clustering with other longbranch orders or with Arachnopulmonata (scorpions and tetrapulmonates). To surmount long-branch attraction, we investigated the effect of taxonomic sampling via sequential deletion of basally branching pseudoscorpion superfamilies, as well as varying gene occupancy thresholds in supermatrices. We show that concatenated supermatrices and coalescent-based summary species tree approaches support a sister group relationship of pseudoscorpions and scorpions, when more of the basally branching taxa are sampled. Matrix completeness had demonstrably less influence on tree topology. As an external arbiter of phylogenetic placement, we leveraged the recent discovery of an ancient genome duplication in the common ancestor of Arachnopulmonata as a litmus test for competing hypotheses of pseudoscorpion relationships. We generated a high-quality developmental transcriptome and the first genome for pseudoscorpions to assess the incidence of arachnopulmonate-specific duplications (e.g., homeobox genes and miRNAs). Our results support the inclusion of pseudoscorpions in Arachnopulmonata (new definition), as the sister group of scorpions. Panscorpiones (new name) is proposed for the clade uniting Scorpiones and Pseudoscorpiones.

Appendix 4

Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida

Ballesteros, J. A., Santibáñez-López, C. E., Baker, C. M., Benavides, L. R., Cunha, T. J., ***Gainett, G.****, Ontano, A. Z., Setton, E. V. W., Arango, C. P., Gavish-Regev, E., Harvey, M. S., Wheeler, W. C., Hormiga, G., Giribet, G., Sharma, P. P.*

Published on 8 February 2022 in Molecular Biology and Evolution

Molecular Biology and Evolution 39 (2), msac021

See attached publication.

Contribution: I contributed transcriptomic data for three Amblypygi species, conducted fieldwork in Israel that sampled scorpions and solifuges used in the analyses and revised the manuscript.

Abstract:

Deciphering the evolutionary relationships of Chelicerata (arachnids, horseshoe crabs, and allied taxa) has proven notoriously difficult, due to their ancient rapid radiation and the incidence of elevated evolutionary rates in several lineages. Although conflicting hypotheses prevail in morphological and molecular data sets alike, the monophyly of Arachnida is nearly universally accepted, despite historical lack of support in molecular data sets. Some phylotranscriptomic analyses have recovered arachnid monophyly, but these did not sample all living orders, whereas analyses including all orders have failed to recover Arachnida. To understand this conflict, we assembled a data set of 506 high-quality genomes and transcriptomes, sampling all living orders of Chelicerata with high occupancy and rigorous approaches to orthology inference. Our analyses consistently recovered the nested placement of horseshoe crabs within a paraphyletic Arachnida. This result was insensitive to variation in evolutionary rates of genes, complexity of the substitution models, and alternative algorithmic approaches to species tree inference. Investigation of sources of systematic bias showed that genes and sites that recover arachnid monophyly are enriched in noise and exhibit low information content. To test the impact of morphological data, we generated a 514-taxon morphological data matrix of extant and fossil Chelicerata, analyzed in tandem with the molecular matrix. Combined analyses recovered the clade Merostomata (the marine orders Xiphosura, Eurypterida, and Chasmataspidida), but merostomates appeared nested within Arachnida. Our results suggest that morphological convergence resulting from adaptations to life in terrestrial habitats has driven the historical perception of arachnid monophyly, paralleling the history of numerous other invertebrate terrestrial groups.

Appendix 5

Recent speciation and phenotypic plasticity within a parthenogenetic lineage of Levantine whip spiders (Chelicerata: Amblypygi: Charinidae)

Baker, C. M., Ballesteros, J. A., Aharon, S., ***Gainett, G.****, Steinpress, I. A., Wizen, G., Sharma, P. P., Gavish-Regev, E.*

Published on 4 July 2022 in Molecular Phylogenetics and Evolution

Molecular Phylogenetics and Evolution 175, 107560.

See attached publication.

Contribution: I conducted fieldwork that sampled the individuals of *Sarax ioanniticus* and *S. israelensis* used in the phylogenetic and species delimitation analyses and revised the manuscript.

Abstract:

Caves constitute ideal study systems for investigating adaptation and speciation, as the abiotic conditions shared by aphotic habitats exert a set of environmental filters on their communities. Arachnids constitute an important component of many cave ecosystems worldwide. We investigated the population genomics of two whip spider species: *Sarax ioanniticus*, a widely distributed parthenogenetic species found across the eastern Mediterranean; and *S. israelensis*, a recently described troglomorphic species that is endemic to caves in Israel. Here, we show that *S. israelensis* is completely genetically distinct from *S. ioanniticus* and most likely also a parthenogen. Counterintuitively, despite the lack of genetic variability within *S. ioanniticus* and *S. israelensis*, we discovered considerable variation in the degree of median eye reduction, particularly in the latter species. Natural history data from captive-bred specimens of *S. israelensis* validated the interpretation of parthenogenesis. Our results are most consistent with a scenario of a sexual ancestral species that underwent speciation, followed by independent transitions to apomictic parthenogenesis in each of the two daughter species. Moreover, the lack of genetic variability suggests that variation in eye morphology in *S. israelensis* is driven exclusively by epigenetic mechanisms.

Appendix 6

Phylogenomics of scorpions reveal contemporaneous diversification of scorpion mammalian predators and mammal-active sodium channel toxins

Santibáñez-López C. E., Aharon, S., Ballesteros, J. A., ***Gainett, G.****, Baker, C. M., González-Santillán, E., Harvey, M. S., Hassan, M. K., Almaaty, A. H. A., Aldeyarbi, S. M., Monod, L., Ojanguren-Affilastro, A., Pinto-da-Rocha, R., Zvik, Y., Gavish-Regev, E., Sharma, P. P.*

Published on 28 March 2022 in Systematic Biology

Systematic Biology 77 (6), 1281–1289.

See attached publication.

Contribution: I conducted field work in Israel that sampled scorpions species used in the analyses and revised the manuscript.

Abstract:

Scorpions constitute a charismatic lineage of arthropods and comprise more than 2500 described species. Found throughout various tropical and temperate habitats, these predatory arachnids have a long evolutionary history, with a fossil record that began in the Silurian. While all scorpions are venomous, the asymmetrically diverse family Buthidae harbors nearly half the diversity of extant scorpions, and all but one of the 58 species that are medically significant to humans. However, the lack of a densely sampled scorpion phylogeny has hindered broader inferences of the diversification dynamics of scorpion toxins. To redress this gap, we assembled a phylogenomic data set of 100 scorpion venom gland transcriptomes and genomes, emphasizing the sampling of highly toxic buthid genera. To infer divergence times of venom gene families, we applied a phylogenomic node dating approach for the species tree in tandem with phylostratigraphic bracketing to estimate the minimum ages of mammal-specific toxins. Our analyses establish a robustly supported phylogeny of scorpions, particularly with regard to relationships between medically significant taxa. Analysis of venom gene families shows that mammal-active sodium channel toxins (NaTx) have independently evolved in five lineages within Buthidae. Temporal windows of mammal-targeting toxin origins are correlated with the basal diversification of major scorpion mammal predators such as shrews, bats, and rodents. These results suggest an evolutionary model of relatively recent diversification of buthid NaTx homologs in response to the diversification of scorpion predators.

Appendix 7

In the land of the blind: Exceptional subterranean speciation of cryptic troglobitic spiders of the genus *Tegenaria* (Araneae: Agelenidae) in Israel

Aharon, S., Ballesteros, J. A., ***Gainett, G.****, Hawlena, D., Sharma, P. P., Gavish-Regev, E.*

Published on 20 April 2023 in Molecular Phylogenetics and Evolution

Molecular Phylogenetics and Evolution 183, 107705

See attached publication.

Contribution: I conducted field work that sampled cave-associated spiders of the genus *Tegenaria* across caves of Israel used in the phylogenetic and species delimitation analyses and revised the manuscript.

Abstract:

Caves have long been recognized as a window into the mechanisms of diversification and convergent evolution, due to the unique conditions of isolation and life in the dark. These lead to adaptations and reduce dispersal and gene flow, resulting in high levels of speciation and endemism. The Israeli cave arachnofauna remains poorly known, but likely represents a rich assemblage. In a recent survey, we found troglophilic funnel-web spiders of the genus *Tegenaria* in 26 caves, present mostly at the cave entrance ecological zone. In addition, we identified at least 14 caves inhabited by troglobitic *Tegenaria,* which are present mostly in the twilight and dark ecological zones. Ten of the caves, located in the north and center of Israel, are inhabited by both troglophilic and troglobitic *Tegenaria*. These spiders bear superficial phenotypic similarities but differ in the levels of eye reduction and pigmentation. To test whether these taxa constitute separate species, as well as understand their relationships to epigean counterparts, we conducted a broad geographic sampling of cave-dwelling *Tegenaria* in Israel and Palestine, using morphological and molecular evidence. Counterintuitively, our results show that the troglobitic *Tegenaria* we studied are distantly related to the troglophilic *Tegenaria* found at each of the cave entrances we sampled. Moreover, seven new troglobitic species can be identified based on genetic differences, eye reduction level, and features of the female and male genitalia. Our COI analysis suggest that the Israeli troglobitic *Tegenaria* species are more closely related to eastern-Mediterranean congeners than to the local sympatric troglophile *Tegenaria* species, suggesting a complex biogeographic history.

Appendix 8

Neglected no longer: Phylogenomic resolution of higher-level relationships in Solifugae

Kulkarni, S. S., Steiner, H. G., Garcia, E. L., Iuri, H., Jones, R. R., Ballesteros, J. A., ***Gainett, G.****, Graham, M.R., Harms, D., Lyle, R., Ojanguren-Affilastro, A.A., Santibañez-López, C.E., Silva de Miranda, G., Cushing, P. E., Gavish-Regev, E., Sharma, P. P.*

Preprinted on October 25, 2022 in bioRxiv

In review in iScience

bioRxiv 2022.10.22.513338

https://doi.org/10.1101/2022.10.22.513338

See attached preprint publication.

Contribution: I conducted fieldwork in Israel sampling solifuges used in the phylogenomic analyses and revised the manuscript.

Abstract:

Considerable progress has been achieved in resolving higher-level relationships of Arthropoda in the past two decades, largely precipitated by advances in sequencing technology. Yet, dark branches persist in the arthropod tree of life, principally among groups that are difficult to collect, occur in cryptic habitats, or are characterized by minute body size. Among chelicerates, the mesodiverse order Solifugae (commonly called camel spiders or sun spiders) is one of the last orders of Arachnida that lacks a higher-level phylogeny altogether and has long been characterized as one of the "neglected cousins", a lineage of arachnid orders that are comparatively poorly studied with respect to evolutionary relationships. Though renowned for their aggression, remarkable running speed, and adaptation to arid habitats, inferring solifuge relationships has been hindered by inaccessibility of diagnostic characters in most ontogenetic stages for morphological datasets, whereas molecular investigations to date have been limited to one of the 12 recognized families. In this study we generated a phylogenomic dataset via capture of ultraconserved elements (UCEs) and sampled all extant families. We recovered a well-resolved phylogeny of solifuge families, with two distinct groups of New World taxa nested within a broader Paleotropical radiation. To provide a temporal context to solifuge diversification, we estimated molecular divergence times using fossil calibrations within a Bayesian framework. Solifugae were inferred to have radiated by the Permian, with divergences of most families dating to the post Paleogene-Cretaceous extinction. These results accord with a diversification history largely driven by vicariance as a result of continental breakup.